VTT Technical Research Centre of Finland

Data typology in manufacturing industries

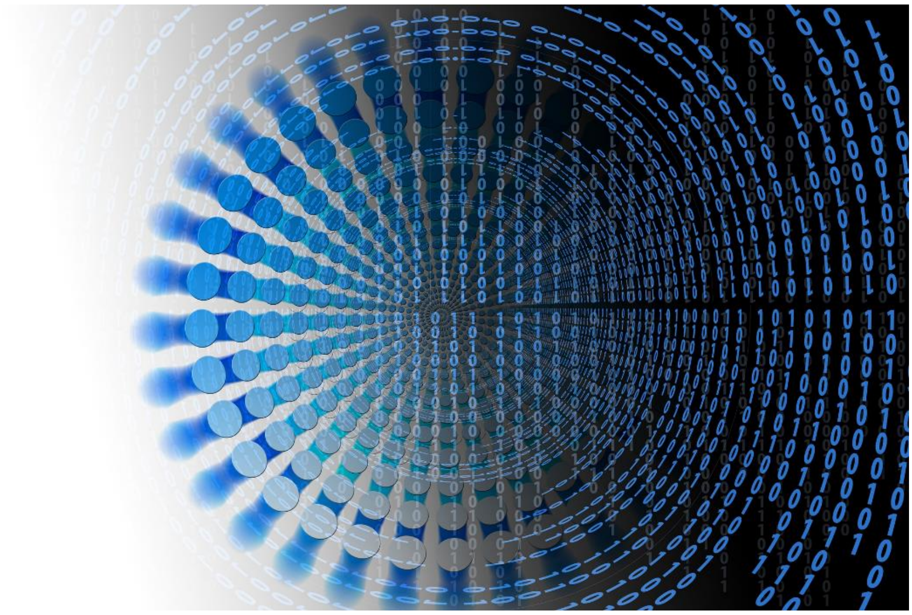
Kortelainen, Helena; Kortelainen, Juha; Pulkkinen, Antti; Juhola, Arto; Hemming, Björn; Kantorovitch, Julia; Heikkilä, Eetu; Ailisto, Heikki; Heilala, Juhani

Published: 29/11/2019

Document Version
Publisher's final version

RESEARCH REPORT

VTT-R-01136-19



Data typology in manufacturing industries

Authors: Helena Kortelainen, Juha Kortelainen, Antti Pulkkinen, Arto Juhola, Björn Hemming, Julia Kantorovitch, Eetu Heikkilä, Heikki Ailisto, Juhani Heilala

Confidentiality: Public

Report's title	
Data typology in manufacturing industries	
Customer, contact person, address VTT Technical Research Centre of Finland Ltd	**Order reference**
Project name Hyper-agile cognitive industry	**Project number/Short name** 121900 / HACI
Author(s) Helena Kortelainen, Juha Kortelainen, Antti Pulkkinen, Arto Juhola, Björn Hemming Julia Kantorovitch, Eetu Heikkilä, Heikki Ailisto, Juhani Heilala	**Pages** 66/2 app.
Keywords data, digitalisation, information, knowledge, typology, engineering, cognitive manufacturing, asset management, services	**Report identification code** VTT-R-01136-19

Summary

Data is becoming the new valuable raw material in business and in our everyday life. Like any other raw material, the value of data increases as the degree of information processing increases. The ongoing hype in artificial intelligence and big data seems to emphasise the amount of data and the assumption often is that the form or structure of data is relatively simple. Data exchange, integration and interoperability are still serious challenges in industry and business, and failing to succeed in meeting these challenges hinders everyday work and leads to productivity losses.

In this report, we focus on defining data as a concept in a comprehensive manner. As the form, structure, amount, source and application of data differs, there is no one definition and no one embodiment of data. The report points out the different forms of dealing with data. We also illustrate the influence of the type, structure and form of data on its use, transformation and exchange, and point out the bottlenecks in application of data, and software development. The report highlights the core domains of the value chain: design and engineering, manufacturing and production, asset life cycle management and life cycle services.

The rapid change of the role of data in industry and business, and especially the new and rising opportunities to create new business and services based on data, have emphasised also the challenges related to data. As for any central resource in business, there are technical, but also many other aspects in using data, such as security, openness and availability, processes and even political issues.

Confidentiality	Public

Espoo 25.11.2019

Written by	Reviewed by	Accepted by
Helena Kortelainen, Principal Scientist	Juhani Heilala, Senior Scientist	Riikka Virkkunen Manager, Digitalising industries

VTT's contact address

VTT Technical Research Centre of Finland Ltd, P.O. Box 1000, FI-02044 VTT, Finland

Distribution (customer and VTT)

VTT Archive, 1 copy

Preface

The meaning and value of data have been evolving fast due to the rapid development of data infrastructure and data analytics, among others. Data is becoming the currency in and valuable raw material for business. In addition, it has increasing importance in industrial and business processes and in our daily lives. The term “data” is present in many places, but there is not much discussion about the nature, form and features of data, i.e. what data actually is in concrete.

In this report, we discuss the term and concept of data from many points of view, including its use, form and structure, life cycle, role in decision-making, quality and security, to mention a few. The goal of this report is to open the different embodiment of data and what needs to be taken into account when using data from different points of view.

The writing of the report and the associated research work has been an interesting journey. Going into essentials of data as a concept and learning its numerous aspects have shown the complexity and challenges in utilising and managing data, but also new opportunities and possibilities.

We thank all the parties that have contributed to the process and have participated in the discussions. The rising importance of data has clearly been recognised and the work for solving the bottlenecks continues.

Espoo 2.12.2019

Authors

Contents

Preface....2
Contents....3
1. Introduction....5
1.1 Background....5
1.2 Purpose of the report....6
1.3 Selected focus domains....7
1.3.1 Design and engineering....7
1.3.2 Manufacturing and production....8
1.3.3 Asset life cycle management and life cycle services....9
2. Characteristics of data....10
2.1 Data structure and information content....10
2.2 Data transfer across asset life cycle phases....12
2.3 The volume of data....13
2.4 The quality and reliability of data....14
2.5 Data ownership....16
2.6 Cyber Security Data....16
2.6.1 Structure of the security data....17
2.6.2 Curation of digital assets....17
3. Defining, modelling, transforming and analysis of data....18
3.1 Defining and modelling data....18
3.1.1 ISO 10303 STEP and data modelling with EXPRESS....18
3.1.2 The Semantic Web....19
3.2 Data exchange, transformation and integration....22
3.3 Cleaning up data....24
3.4 Creating value from data....24
3.4.1 Data value creation process....24
3.4.2 Data exploration and descriptive data analysis....25
3.4.3 Predictive data modelling....25
3.4.4 Supporting business decisions....26
4. Industrial Internet of Things....26
5. Artificial intelligence and data analytics....28
5.1.1 Data transformation using AI....31
6. Challenges and opportunities in using data....31
6.1 Design and engineering....31
6.2 Manufacturing....31
6.2.1 Product Lifecycle Management: definitions, structures, variety and closed-loop....32
6.2.2 Cyber-physical systems and manufacturing, architecture and 4.0....36
6.3 Supply chain management....37
6.4 Production activities functional model and data flows....39
6.5 Asset management and life cycle services....43
7. Data interoperability and standardisation....45
7.1 Reference Architectures for Interoperable Manufacturing....45
7.2 Standardisation for smart manufacturing....47

8. Summary and future research topics....49

8.1 Summary....49
8.2 Future research topics....50
8.2.1 Vendor locks....50
8.2.2 Data as a digital asset....50
8.2.3 Refining and sharing data in an ecosystem....50
8.2.4 The lifetime of the data....50
8.2.5 Safety and security of AI in manufacturing....51
8.2.6 Automated data interoperability....51
8.2.7 Knowledge management as a business asset....51
8.2.8 Accuracy, reliability and traceability of the data....52
8.2.9 Relating customisation and agility to economy for decision-making....52
8.2.10 Data in circular manufacturing....52

References....53

9. Appendices....62

APPEDIX 1: Abbreviations and terminology....62
APPENDIX 2: Overview of data required for data security (not a comprehensive list)....65

1. Introduction

1.1 Background

Data is becoming the new valuable raw material in business and in our everyday life. Like any other raw material, the value of data increases as the degree of information processing increases. In addition, the amount of available data often affects its value, as data analysis requires a sufficient amount of representative data. The term *data* is often used carelessly and it gives an impression that the term is explicitly defined and it has a relatively narrow scope. The ongoing hype in artificial intelligence (AI) and big data seems to emphasise the amount of data and the assumption often is that the form or structure of data is relatively simple. Data exchange, integration and interoperability are still remarkable problems in industry and in business, and they hinder everyday work and cause loss of productivity.

Data is also in the core of Industry 4.0 and in the digitalisation process in general. The concept of Digital twin is used to describe the data content that mirrors the physical entity. NASA defined Digital twin (DT) in 2010 as “an integrated multi-physics, multi-scale, probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its flying twin (Shafto et al., 2010). Since that the concept of DT was extended to embrace other aspects like the life cycle view (Tuegel, 2012) and prognostics and diagnostics activities (Reifsnider et al., 2013) and manufacturing (Helu et al., 2017, Hedberg et al., 2016). Lee and colleagues (Lee et al., 2013) further extended the DT concept beyond individual products and regarded DT as a virtual counterpart of production resources in smart manufacturing.

Generally, life cycles of physical items (e.g. products, assets or production systems) or services entail phases such as concept, development, realisation, utilisation, enhancement, retirement, reuse and recycling (e.g. IEC, 2014, ISO/IEC/IEEE 1528) and the life cycle stages are presented as a linear chain (Figure *1*). Industrial systems and complex products consist of subsystems, items and components – all of which have a life cycle of their own.

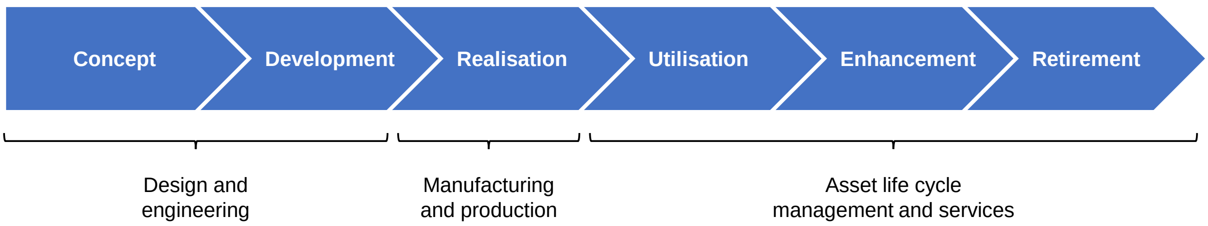


Figure 1: Life cycle stages in asset, product or service life.

Management of the data in current manufacturing and production systems over the life cycle of individual products and assets and across manufacturing enterprise processes and networks is a challenging task. Further challenges to life cycle data management are posed by the requirements of sustainability and circular economy that emphasise cradle-to grave approaches (Seliger, 2007). In all the life cycle stages as presented in Figure *1*, core tasks include collecting, handling and integration of data from various, heterogeneous sources and planning for the use of data in effective design and planning, in efficient manufacturing and supplier chain management, and in optimised execution and improvement of operations. Data must be of high quality, secure and real-time for data-driven decision-making.

Industrial companies handle rapidly increasing amounts of data. As an example, Bernard Marr from Forbes[1] evaluated that “The amount of data we produce every day is truly mind-boggling.

[1] https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#6dcc2a9360ba

There are 2.5 quintillion bytes of data created each day at our current pace, but that pace is only accelerating with the growth of the Internet of Things (IoT)".

Figure 2. The coming flood of data (Meariam, 2017)

Figure 2 illustrates the future data flood that autonomous vehicles will produce. Even though the amount of the data created is exponentially increasing along with the means to produce data, big data analytics, AI and analytics are not new. An example from Shell Expro was given by John Woodhouse (2018): in the 1990s, Shell analysed the 20 years of combined history of all North Sea assets. There was 10^8 work orders in a single database and two years was used for data cleaning. As a result four cases (approx 3%) showed statistically significant relationships (non-randomness). However, solving one case resulted in benefits that paid for the entire effort.

Smart manufacturing relies on the evolution of cognitive capabilities in manufacturing including development of AI-enabled robots that can collaborate closely with humans and learn their behaviour, and even cognitive systems that can learn with experience, solve problems and make decisions without human intervention. Cognitive machines and systems create enormous amounts of data. Cognitive capabilities are needed to convert the exploding big data to meaningful insights that further improve manufacturing processes and functions (Frost and Sullivan, 2017).

Emerging trends in products, processes, materials and technologies will transform the manufacturing industry landscape. An inevitable shift to leaner, smarter, more flexible production will have a significant impact on factory design, operation and control, supply chains and the nature of work. The increased digitalisation and the use of opportunities arising from data collection and analysis will allow more rapid, responsive and more connected manufacturing where products and production processes co-develop to adapt to the changing customer demands, markets, and the conditions in factories and enterprises (ElMaraghy, 2019).

1.2 Purpose of the report

In 2019, VTT launched the Hyper Agile Cognitive Industry initiative that aims at seamless integration of information technology systems across the manufacturing value chain. Hyper Agile Cognitive Manufacturing initiative focuses on integrating and using data from design, production, and service life in decision making throughout the product life cycle. In this report, we focus on defining data as a concept in detail and in a comprehensive manner. As the form or structure, amount, source and application of data differs, there is no one definition and no one embodiment of data. One of the goals for this report is to point out these different forms of dealing with data. In addition, the influence of the type, structure and form of data on its transformation and exchange is illustrated and the bottlenecks in application of data and information, and in the development of software tools and applications, are pointed out. We narrow the discussion to data that is in machine-readable format, i.e. it can be stored, processed, manipulated and utilised with software applications in computer systems.

The scope of the report is in the manufacturing industry, covering the whole product or system life cycle – from the pre-concept design phase, through design and engineering, manufacturing or production, use or operation together with maintenance and support services, until the end-of-life of the product or system. Three focus domains are selected for examples: 1) engineering and design, 2) manufacturing and production, and 3) asset management and asset life cycle services. The connecting elements over life cycle phases are the system-level approach and the related data. This report aims at creating common ground for understanding the different dimensions of data and at contributing to the discussion on data as an asset.

1.3 Selected focus domains

The smart manufacturing industry will be agile and capable of adapting to a wide variety of changing conditions like customer needs, production uncertainty, and market changes. Such cognitive characteristics require data. Smart manufacturing is not only about the manufacturing process itself. It requires not only horizontal and vertical integration, but also End-to-End Integration that addresses the entire product´s value chain across different companies (Kagermann et al., 2013). This report highlights the value chain approach and concentrates on three core domains: design and engineering, manufacturing and production and asset life cycle management and life cycle services.

1.3.1 Design and engineering

In technical design and engineering, i.e. engineering design, the primary process and target is to design new products, systems and services. During the process, a large amount of data in a different level of detail and in a different form is typically produced, modified and managed. The data is primarily valuable for the design process but often has value during the later product or system life cycle phases. An example of this is plant design. During the design and engineering of a process plant, a large amount of data is produced concerning, e.g. the selection of the process equipment, such as pumps, valves, and measurement and control equipment. The data contains information about the shape, type, location, relations and other design details of the components, equipment and systems. In addition, it may contain, e.g. simulation and analysis data about the dynamics of the process. This data is usually not produced for the operation of the plant and is not necessarily included in the handover of the plant data from the engineering service provider to the plant operator. Instead, a cleaned up set of data containing the necessary outcome of the design and engineering for operating the plant is provided to the operator. The same kind of approach applies in machine design in mechanical engineering, except that even a cleaned up set of product design data is rarely provided to the operator or owner of the product. Another similar example is manufacturing and production data. The data is necessary and valuable for designing, ramping up and for the operation of the manufacturing facility, but does not provide added value for the later product or system life cycle phases. In addition, it typically contains business critical information about the manufacturer and should not be transferred to third parties.

For the design phase data, one characteristic feature is the complexity of the data. Complexity in this context means that the form and information content of the data is rich and using the data requires dedicated software tools. In addition, the data formats and content are usually defined by the software vendors and the data models and definitions are not necessarily available to the users. This often makes the integration and reuse of existing design data challenging.

The increasing emphasis on providing digital services supporting physical products or systems is widening the useful scope of data over the product or system life cycle. For example, the digital design information about a process plant or a machine system can be utilised in the digital support services, such as condition monitoring and maintenance. Digital 3D design models of a plant or simulation models of a machine can be the basis for digital twins or visualisation of data analytics for maintenance. The digital design data is often needed in problem solving during the operation of the product or a system, or when the system is upgraded. Especially in the latter case, the compatibility of the data can become an issue due to continuously changing data formats.

1.3.2 Manufacturing and production

Apart from production orders, the execution of manufacturing operations depends on engineering design and manufacturing capabilities, such as resources and product definition. The product definition *per se* is hardly ever ready for manufacturing purposes, but the definition requires further elaborations and additions or even corrections. The definition is interpreted and adjusted to meet the capabilities at hand. For example, the functional tolerances of part of the definition have to be translated into manufacturing tolerancing systems. The former is defined by product design engineering and the latter requires manufacturing knowledge and is seldom available for designers.

Manufacturing planning and operations use and produce many kinds of datasets. Manufacturing product and process design-related data involves all the data for supporting the functions and processes from conceiving and developing new (and improved) products and manufacturing processes to start manufacturing execution. Manufacturing processes encompass all of the functions associated with translating product designs into finished goods. Manufacturing infrastructure encompasses all of the functions that support the creation of product, both directly and indirectly. The infrastructure and the skills of the resources are in a key role for defining manufacturing capabilities.

The enterprise management point of view covers all of the functions associated with managing the operation and maintenance of a manufacturing business entity. Therefore, the production orders and manufacturing capabilities as well as the maintenance of the capabilities are essential for enterprise management. However, the control and development of a manufacturing business and its technology environment encompasses the strategic changes of the manufacturing entity in an enterprise.

There are five main reasons why manufacturers collect, transfer, store and analyse data for their production process:

i. to conduct R&D and testing in the pre-production phase,
ii. to manage actual part manufacturing, internal logistics and final assembly,
iii. for overarching control and coordination of the geographically spread-out production,
iv. for efficient supply chain management and the smooth flow of goods, services, and capital necessary for production,
v. for the after-sales and product life cycle service business.

The execution of manufacturing operations and the control of the operations are separated to different organisations. As said, manufacturing operations are also dependant on the other organisational entities, such as sales, business management and engineering design. Thus, the data sets are spread out to different organisational entities and stored in the data management systems of the entities. Processes and value creation necessitates the integration in many directions:

- *Horizontal integration through value networks*: integration of the various IT systems used in the different stages of the design and engineering, manufacturing and business planning processes within a company (e.g. inbound logistics, production, outbound logistics, marketing) and between several different companies (value networks).

- *Vertical integration and networked manufacturing systems*: integration of the various IT systems at the different hierarchical levels (e.g. actuator and sensor level, manufacturing and execution level, production management level, and corporate planning levels) to deliver an end-to-end solution.
- *End-to-end digital integration across the entire value network*: integration throughout the engineering process so that the digital and real worlds are integrated across a product's entire value chain and across different companies, whilst also incorporating customer requirements.

A brief summary of each integration feature and each priority area for action is provided by Kagermann et al. (2013).

1.3.3 Asset life cycle management and life cycle services

Asset management is the set of coordinated activities that an organisation uses to realise value from assets in the delivery of its objectives. Asset management aims at balancing costs, risks and benefits, often over different timescales. An asset life cycle management plan is the tool to implement asset management activities (Hastings, 2015). Such a plan is formed by identifying the operating regime of the asset, the associated maintenance, repair and improvement activities, the planned life, and the disposal plan. The life cycle management plan is then a basis for planning resources, budgets, replacements, and upgrades.

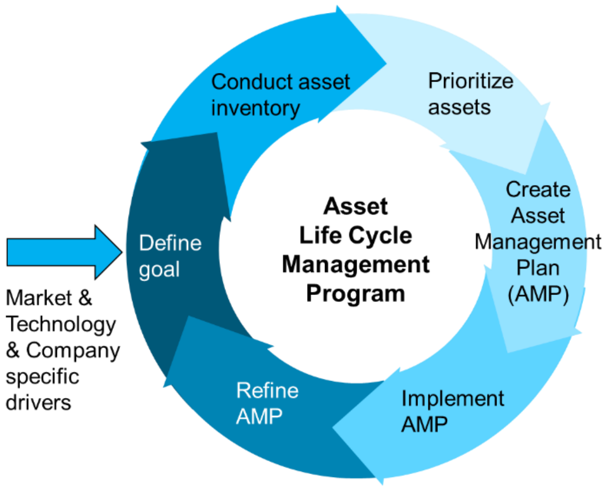


Figure 3. Asset life cycle management programme.

Asset management standard ISO 55000 calls also for a systematic approach for determining the information needs related to the assets and asset management. The required information deals with technical and physical asset properties, service delivery and operations, maintenance management, performance management and reporting, financial and resource management and contract management. As companies outsource their activities and acquire life cycle services from external companies, standards like ISO 55000 and EN17007 provide a common basis for supplier-customer collaboration (Kortelainen & Komonen, 2017).

Novel advancements of IoT enable the machine and service suppliers access to information at customer sites. Access to the data opens new possibilities for typical fleet services like maintenance and also for new business models, e.g. performance-based contracts and shared use of assets, 3rd party services and even novel digital product lines. However, the applicability of the data or the development of services is not straightforward as the installed products differ from each other (e.g. age, type, performance, capacity), with different environments and operational modes (Kortelainen et al., 2017).

2. Characteristics of data

The data, information, knowledge and wisdom (DIKW) hierarchy is widely accepted as a basic model describing levels of understanding of issues in consideration. Ackoff (1989) was one of the early authors of this model, and since the model has been referred to in the literature as the 'Knowledge hierarchy', 'Information hierarchy' and the 'Knowledge pyramid' (see Figure *4*).

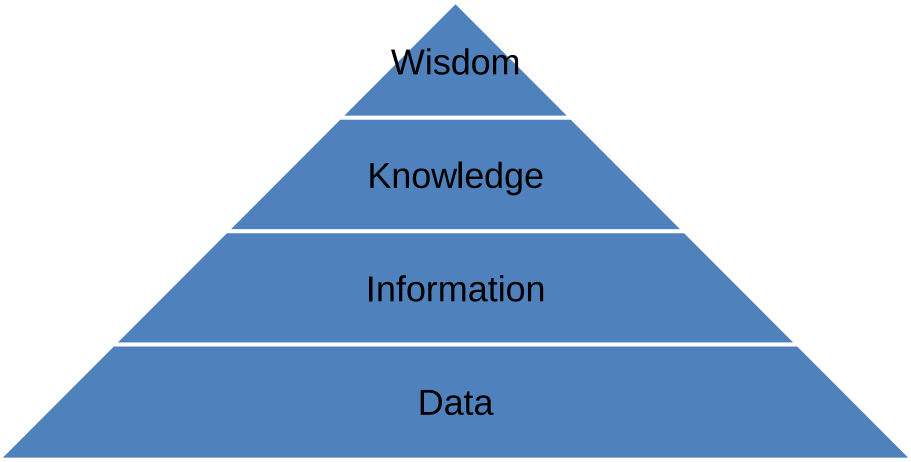


Figure 4: The data-information-knowledge-wisdom pyramid.

In the model presented by Ackoff (1989) and Rowley (2006), data refers to the symbols that represent the properties of objects and events and they are the products of observation. Information consists of processed data, with the processing directed at increasing its usefulness. The difference between data and information is functional, not structural. Information is contained in descriptions, answers to questions that begin with such words as who, what, when, where, and how many. Information systems generate, store, retrieve and process data. Information is derived from data, and data becomes information when it has some meaning and value to its user. Knowledge is conveyed by instructions, answers to how-to questions. Understanding is conveyed by explanations, answers to why questions. Knowledge can be obtained either by transmission from another who has it, by instruction, or by extracting it from experience.

The DIKW hierarchy also includes the concept of wisdom. Wisdom deals with values. It involves the exercise of judgment. Evaluations of efficiency are all based on a logic that, in principle, can be programmed into a computer and automated. These evaluative principles are impersonal. "*Intelligence is the ability to increase efficiency, wisdom is the ability to increase effectiveness*" (Ackoff, 1999).

Each step up the knowledge pyramid adds value to the data as the data is enriched with meaning and context. In this process, the data is refined to the knowledge and insights that allow making interpretations and applying the derived knowledge beyond the original data collection scheme. The DIKW hierarchy and the process of refining the data with experiences to guide actions is illustrated in a practical example given in Figure *33*.

The DIKW hierarchy is by far not the only model to categorise different data types. Other concepts related to data are collected in Appendix 1. Abbreviations and terminology.

2.1 Data structure and information content

The fast development of the Internet of Things (IoT) and its industrial variant, the Industrial Internet of Things (IIoT), has emphasised the value of a large amount of measured or collected data and its utilisation in different processes. The IoT data is often by nature a data stream that comes from defined sources and is exploited instantly. The data is used for monitoring or some real-time analytics are computed based on it. The IoT system may contain also static or relatively static data, such as design information. The design and engineering data is typically typed, structured and complex, i.e. it has several data types and different levels of meaning (low level types, such as integers and strings, higher level types, such as machine equipment types, etc.). Complex data structures and rich data types often cause problems in data interoperability.

A data structure indicates how the data is organised and formatted to make it easier to use, access and manage. One way to assimilate data structure is to compare it to a book. In a text book, the text is divided, e.g., into chapters and sections, and the body text in sections is divided into paragraphs and sentences. Books often have a table of contents, which gives the reader a convenient means to look for particular topics in the book, without having to read it throughout. Data structures are used in a similar manner to ease and improve the efficiency of using the data, but usually from the software application or system point of view. Below is an example[2] of structuring data with Extensible Markup Language, XML (World Wide Web Consortium, 2016; World Wide Web Consortium, 2006):

```
<?xml version="1.0" encoding="UTF-8"?>
<breakfast_menu>
<food>
    <name>Belgian Waffles</name>
    <price>$5.95</price>
    <description>
   Two of our famous Belgian Waffles with plenty of real maple syrup
   </description>
    <calories>650</calories>
</food>
<food>
    <name>Strawberry Belgian Waffles</name>
    <price>$7.95</price>
    <description>
    Light Belgian waffles covered with strawberries and whipped cream
    </description>
    <calories>900</calories>
</food>
<food>
    <name>Berry-Berry Belgian Waffles</name>
    <price>$8.95</price>
    <description>
    Belgian waffles covered with assorted fresh berries and whipped cream
    </description>
    <calories>900</calories>
</food>
<food>
    <name>French Toast</name>
    <price>$4.50</price>
    <description>
    Thick slices made from our homemade sourdough bread
    </description>
    <calories>600</calories>
</food>
<food>
    <name>Homestyle Breakfast</name>
    <price>$6.95</price>
    <description>
    Two eggs, bacon or sausage, toast, and our ever-popular hash browns
    </description>
    <calories>950</calories>
</food>
</breakfast_menu>
```

[2] W3Schools XML Tutorial, XML Example 2: https://www.w3schools.com/xml/

Here, the information of a breakfast menu is structured with additional XML tags in the data, such as a tag pair ”<food>” and “</food>”. The tags in the above example defines the structural elements of the data but do not add any particular content to it. The structural elements can be used, e.g., for formatting the data for the reader or to make it easier for software applications to interpret the data and, e.g., find how many different dishes the menu contains. In the above example, the efficient use of the data requires that the format and the semantics of the data structuring mechanism are available for the data users, or more precisely to the software developers of the software applications used by the data users. XML is a good example of such a mechanism and its standardisation. Data structures can be very large and complex to cover all the needs of all the data stakeholders.

Databases (DB) and database management systems (DBMS) provide means and tools for structuring and managing large-scale and complex data. The so-called relational databases (RDB) structure the data into tables with rows and columns of data elements of different types. The table represent a collection of entities of one type, the rows of the table represent instances of the entity type and the columns of the table represent the attributes of the instances. In the table, each instance of the entity type (i.e. each row) has a unique identifier (see Figure 5).

ID	*First name*	*Surname*	*Unit*	*Phone number*	*Email address*
1	John	Doe	AB987	+358 20 123 4567	john.doe@acmecorp.com
2	Matti	Meikäläinen	AB123	+358 20 123 5678	matti.meikalainen@acmecorp.com
3	Fun	Doo	AC777	+358 20 123 6789	fun.doo@acmecorp.com
…	…	…	…	…	…

Figure 5: Example of a relational database structure. The table represent a collection of contact information of people from one organisation. The header line represents the attribute types for each instance (person) and the other lines represent the contact information of one person.

As a relational database management system (RDBMS) may be implemented in numerous ways, there are no guarantees that the data in such an implementation can be accessed through a common interface. To simplify the use of RDBMSs, some standardisation has been done to create a common language to define the collaboration with the RDBMSs, i.e. Structured Query Language, SQL (ISO/IEC 9075-1:2016, 2016). On the other hand, an example of large and complex data structures and their standardisation is the set of MIMOSA open standards for physical asset management (MIMOSA, 2019).

As data in general and digital services in particular are becoming increasingly common related to both consumers as well as in industrial products, the data interoperability – or the lack of it – is becoming an issue. Standardisation has been one of the solutions to improve data interoperability, but the slow process in standardisation and challenges in funding the standardisation work are are shifting focus to other possible complementary solutions. One of the interesting directions is data modelling and the ability to automatically map the data from one system to another, i.e. automated data transformation. Also in this approach, some standardisation is needed to define the mechanisms to define the source and the target data, and how the transformation is done, especially when the transformation is not explicit.

2.2 Data transfer across asset life cycle phases

Data is produced, acquired and used throughout the product, asset or system life cycle (see Figure 1). During the concept and design phase, domain and process knowledge are generated based on functionality and formal specifications, and stored in the system’s knowledge base (KB). During post deployment – i.e., the utilisation and enhancement phase – context- and user-specific knowledge are created and corresponding data is collected in the plant information systems (e.g. ERP, CMMS, process automation and control system). This dictates that the design must accommodate re-evaluation and update of the knowledge.

Digital twin is a concept that interlinks and ties together the data life cycle with that of a physical item. A digital twin is the representation of a physical asset in the digital world (Lee et al., 2015). As illustrated in Figure 6, a digital twin is generated and maintained as synchronised with the evolving behaviour of the physical asset, thus allowing simulations, functional analytics and predictive and prescriptive analytics with real-time data.

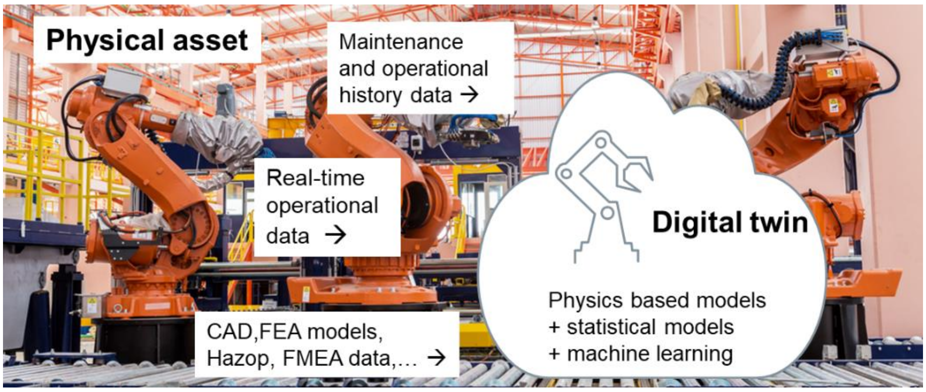


Figure 6: The Digital Twin concept.

Knowledge-based systems are always designed with tasks in mind (Chandrasekaran, Johnson, & Smith, 1992). Therefore, the relation between tasks/processes, knowledge and mechanisms supporting knowledge sharing & transfer should by carefully analysed during the design phase. This can be achieved, for example, by the use of ontologies that include information on the domains in which a smart products system is expected to be used, the processes it is designed to support and the attributes or features exposed. In addition, the data items specifying the knowledge that a smart product requires, user/relevant stakeholders' characteristics, and the knowledge on the smart product interaction with its environment and information exchange with other products and their users (Sabou & Kantorovitch, 2009).

2.3 The volume of data

Big data describes such datasets which could not be acquired, stored, and managed by classic database software (Chen et al., 2014). It is typical for big data that there is a huge amount of data and it is in different forms (structured, semi-structured, and unstructured). Big data is classified with the following characteristics (Chen et al., 2014):

1) Volume (great volume),
2) Variety (various modalities),
3) Velocity (rapid generation), and
4) Value (huge value but very low density)

The volume of data, and especially the increase of the speed at which data is produced, has been under active debate during the past couple of years. For example, Bernard Marr wrote in his article in Forbes (Marr, 2018) that 2.5 quintillion bytes of data is created each day (in 2018) and over 90% of the existing data has been produced over the last two years (again, in 2018). It is important to notice that this concerns the produced data, not information. The produced data includes, among other things, search requests in Google and other search services, social media activities, Internet-based teleconferencing calls, video and audio streaming, and the data that goes through IoT.

Companies tend to produce increasing numbers of product designs, which eventually leads to large volumes of product data to be managed. In 2008 a company, which we have had long-term collaboration with, indicated that it had 1.4 million items, 2 million item structure rows and 750 000 documents in the Product Data Management (PDM) system. During one afternoon hour in late August, 2008, 143 users made more than a thousand operations in the system. They created 73 new documents and 17 new items in the system. Also, the users sought and checked the data of 440 documents and 725 items, opened nearly 300 document files and printed 125 bills of materials. After four years, in 2012, the company had 1.7 million items, 2.6 item structure rows and 800 000 documents in the system. Thus, there was an increase of items by 21%, structure rows by 30% and documents by 7%, in just 4 years. The increase actually took place during the recovery from the economic recession, when the introduction of new products and engineering design activities was not very intense, in general. Naturally, this amount of data is not a problem for databases or PDM systems, but also the data-related operations and transactions can easily rise and require a lot of engineering hours (see Figure 7), if data management is not automatic.

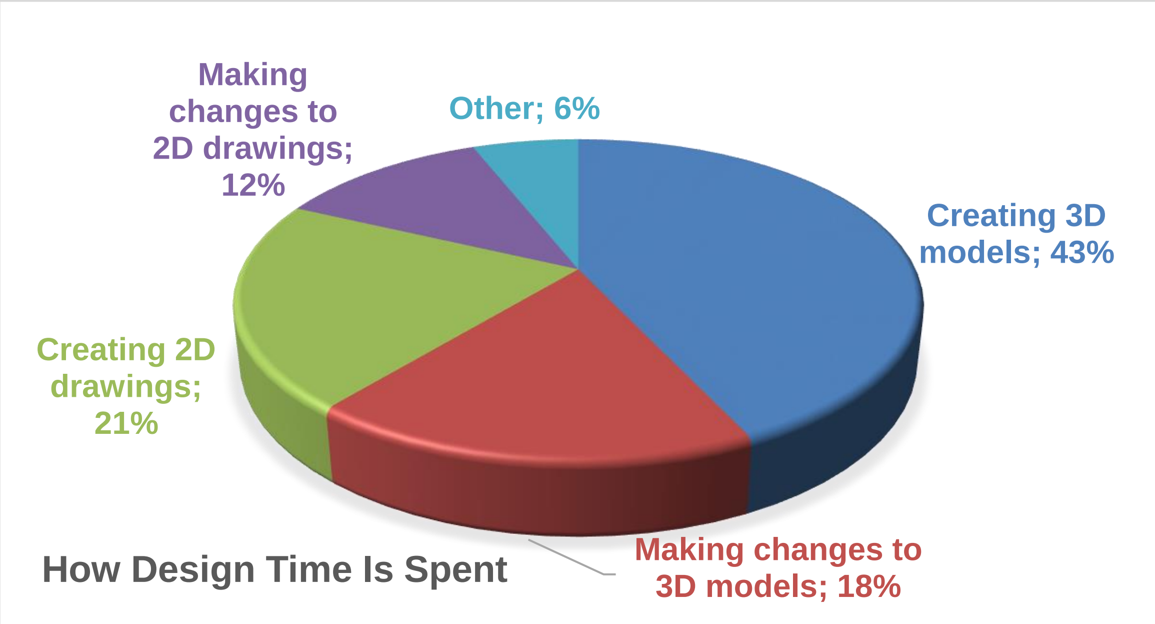


Figure 7: An approximation on the designers' work categories (source: https://tech-clarity.com/adopting-model-based-definition-mbd/6343).

In other industrial contexts, the volume of data shows the same kind of trend as discussed above. The wide application of the Industrial Internet is raising the production and utilisation of data to a new level. At the same time, the methods and tools for analysing, integrating and utilising data are becoming more sophisticated and efficient, i.e., it is possible to turn the data into valuable information. The fast development of Artificial Intelligence and all the related technologies are opening new possibilities also in the industrial context.

2.4 The quality and reliability of data

Dataset quality is crucial for the potential to extract information from the data. In general, the quality of a decision is hardly better than the quality of data used to support the decision. On the other hand, outliers of data from a sensor network may be detected and deleted by statistical reasoning or by an algorithm in artificial intelligence. Still, at least in the first steps of digitalisation, bad data quality may result in significant losses due to interruptions in production or bad-quality of products.

Cai (2015) suggested that dimensions of data quality are availability, usability, reliability, relevance and presentation quality. Usability entails also credibility. Reliability can be split into elements like accuracy, consistency, integrity and completeness. At this level, also different interpretations exist. For instance, credibility is regarded as a part of accuracy because accuracy cannot be proved without documents and verifications – which form credibility. However, the dimensions presented by Cai (2015) have proved very useful in the data assessment process.

The aspects of data quality are also incorporated in the information quality framework by Eppler (2006), see Figure 8.

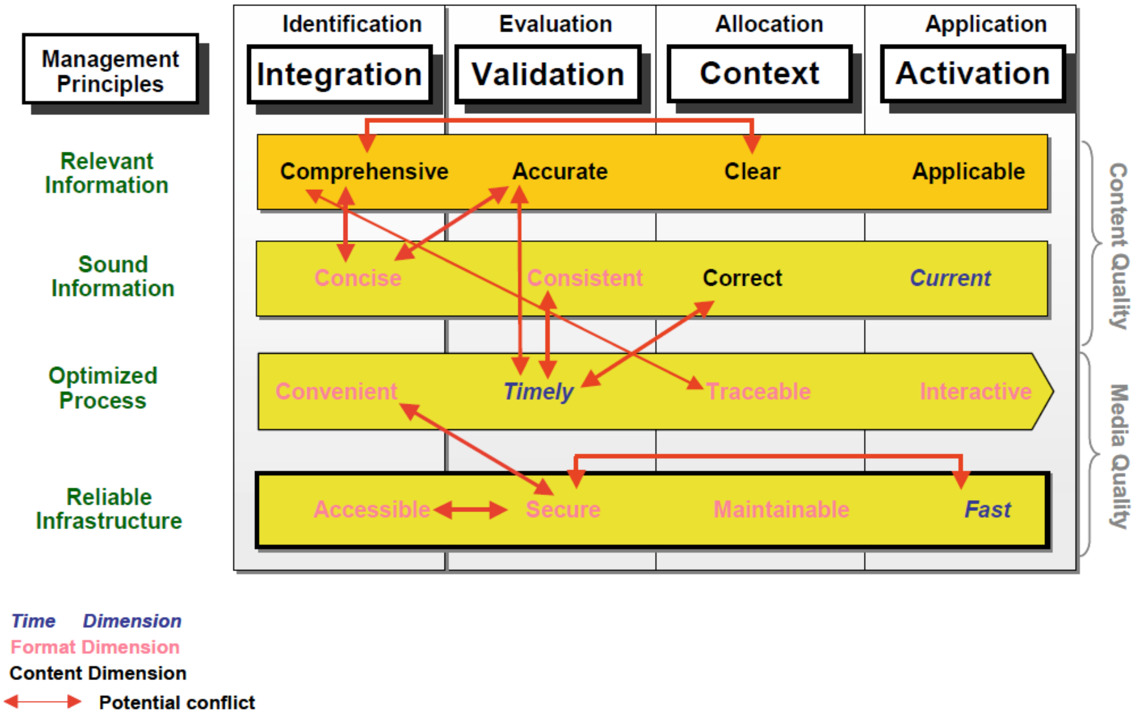


Figure 8. The information quality framework according to Eppler (2006) p. 68.

According to the framework, the quality of information depends on the timeliness, format and content of the information. Also, the media of the data affects data quality. How comprehensive, accurate, clear and applicable the content is defines how relevant the content of information is. How concise, consistent, correct and current the information is, dictates how sound the information content is (see Figure 8). In an optimised information delivery and management process, there should be no issues related to convenience, timelines, traceability or interactivity of the information. The reliability of the information is related to the accessibility, security, maintainability and speed of the media. The mentioned aspects are, however, dependent on each other. For example, the consistency problem related to interrelated knowledge bases or data sets in different databases is a well-known issue. It can be demonstrated with the set-up of two clocks that share the same data, i.e. the instance of time, but are not related to each other nor to a reference source of data. The indicated data is hardly the same and the information of the two sources is inconsistent. In this case the set up and media quality have an effect on the content quality. The suggest approach to alleviate the consistency problem is to favour a single-source approach, such as an integrated platform for data management.

When decisions are made based on acquired data, it is helpful if information of reliability or even an uncertainty range is available. Ideally, every piece of acquired data should be accompanied by an uncertainty range. The highest level of DIKW hierarchy (see section 1.2) is Wisdom, based on which decisions can be made. The example of icy roads as the temperature is close to zero degrees Celsius can be expanded by variation of temperature because of uncertainty of data and variation of temperature due to location. The result is then probability and occurrence of an icy road. Transferred to manufacturing, a similar example for data reliability could be: “at a confidence level of 99% it is concluded that 5% of produced components are not within specifications”. When quality and data quality are described in numbers, improvements and optimisation becomes possible.

2.5 Data ownership

Data ownership is an important aspect when it comes to offering data and negotiating contracts in a digital business ecosystem. One possible means of giving users control of their data would be the allocation of property rights over machine-generated data to the data generators. This could be considered a means to facilitate trade and the creation of data markets. In addition, it could respond to a perceived need to ensure a fairer distribution of value across the data value chain. Currently, data services and data users are able to drive great profits from machine-generated data without sharing them with the data generators (EPSC, 2017).

The development of digital technologies creates a need to develop a common understanding of ownership rights regarding the data created by connected devices. A sensor manufactured by one company can operate in a system developed by another, and be deployed in an environment owned by a third. Agreement will be needed on who has which rights to the resulting data (OECD 2017). Also BDVA (2019) emphasises that data sharing space will only materialise if data producers are guaranteed to retain their rights as the original owners. The Industrial Data Space is one approach to address data ownership by providing a secure and trusted platform for authorisation and authentication within a decentralised architecture (International Data Spaces Association, 2018).

However, the creation of ownership rights for data would entail a number of risks (EPSC, 2017).

- It may lower companies' incentives to invest in data analytics as their share of the value creation decreases.
- There would be significant pragmatic challenges. The biggest part of the value is generated through aggregation. Therefore, it would be difficult to craft a rule to 'fairly' assign part of that value to the single generators of the data.
- The more the data economy evolves, the more the bulk of the value will tend to be generated in the 'last mile' of the big data value chain.
- Introduction of data property rights would incur high implementation costs for single individuals and small to medium enterprises (SMEs) wishing to enforce them and would necessarily involve transaction costs that would reduce the incentive to trade the data.

Information and data may belong to various actors, but they cannot be owned in the legislative sense. Information can be managed. The owner of a device or service is usually able to prevent others from accessing the data by preventing access to the device or service. It is possible to specify in contracts, e.g. who data belongs to, and what access rights there are to the data. A limitation of contracts is that they cannot be binding on a third party (Seppälä et al., 2018). Data sharing has been shown to increase the collaboration and informal dialogue between companies. Data should not be considered only from the viewpoint of return of investment but also regarding the amount of reuse when it is compared to the effort of publishing; the value of data derives from use and reuse instead of ownership (Kortelainen et al., 2017).

2.6 Cyber Security Data

Traditionally, digital data security has meant protecting data from unauthorised disclosure, corruption or wipe-out. The respective remedies have been various crypto and obfuscation algorithms, access controls, fault-tolerance techniques and back-up strategies.

All these methods generate auxiliary data, multiple copies or especially checking/error recovery data that must be likewise stored, cared for and managed. The auxiliary data includes various logs, inventories, directories, certificates, keys, policies, and (security) control mechanism configurations and specifically generated higher-level security management data like emergency plans and system state summaries. None of these are static; updates, and possible (co-operative) intelligence gathering processes are to be arranged and managed as well.

We limit ourselves in this section to considering this auxiliary data, not the data to be protected.

2.6.1 Structure of the security data

As for the structuring of the security data, there are diverse standards, but for example with logging, there are quite a many of them, and not too detailed, leading to application specifics and tedious parsing for a unified view.

To make sense of the overall situation, complex SIEM – security information and event management systems are a possibility in high-risk environments. These tend to have taxonomies of their own, both on analysis – with specialist data fusion algorithms for refining raw data and distilling anomalous events – and a (policy-based) management side of things. There are, however, some widely accepted methodologies and repositories with associated taxonomies.

For classification of stages of attacks there is, e.g., the one outlined by Richard Bejtlich in his book "*The Tao of network security monitoring : beyond intrusion detection*" (Bejtlich, 2005). For analysing different threat categories, see Shostack (2014), MITRE ATT&CK (2019) and NIST (2012). For risk evaluations, a commendable methodology is FAIR – Factor Analysis of Information Risk (Freund & Jones, 2014). FAIR divides the risk into its constituent parts. Another is DREAD (Shostack, 2008), but this is not currently favoured, since it has been applied to threat analysis where it is not very fitting. It can be rather described as a simplified risk analysis method.

A crucial step in estimating risks is the probabilities of threat events. This is not a straightforward task. To start with, these depend on specific systems, configurations and circumstances. Frequencies of historical incidents are not very usable due to their (in most cases) infrequency. Some mathematical approaches could be used like Monte Carlo, statistical analysis for confidence intervals, fuzzy logic and the like, but in the end, it might boil down to "expert opinion" and security intelligence findings, projected to the case at hand. Nevertheless, data about past security events, with related system composition, configuration, known vulnerabilities and state is maintained.

Efforts on systemising the management and evaluation of information security – with appropriate definitions and models – include the ISO 27000 series of standards and ISO/IEC 15408 Common Criteria for Information Technology Security Evaluation. On the research front, Savola (2009) and Savola et al. (2012) have worked on security metrics taxonomies.

2.6.2 Curation of digital assets

Proper curation of digital assets is critical to daily operations and requires permanent access and maintenance of trustworthy data. Corruption of that data can have catastrophic consequences on product development and affect viability of an enterprise. There is a need to protect the product data and its owner(s) by providing authorisation, authentication, and traceability of trustworthy product data through the product life cycle. Currently, knowing who can use data, how the data can be used, and who did what to the data is mainly captured in contracts and manual paper-based tracking methods. Industry needs a faster, more secure and sustainable way to record, embed or link authentication, authorisation, and traceability information to the product data (NIST, 2018).

3. Defining, modelling, transforming and analysis of data

As already discussed in the Introduction section, data is not one homogeneous mass that can be utilised in a similar manner as, e.g., electricity. The structure, representation, meaning and semantics can differ, which make the utilisation of data sometimes challenging. When the same data is utilised by many parties and possibly with several different software applications, these data features become crucial for fluent work flows.

This section addresses two views on the modelling of data. First, modelling of data itself and, secondly, using the data to model different concepts.

3.1 Defining and modelling data

The increasing use of data in industry and business, and the dependency of processes and services on data, set demands on data applicability. The meaning, validity and form of the data need to be known for its reliable application. In addition, the more data is involved in processes, the more fluent its application needs to be. For complex data, data modelling, i.e. defining the meaning, form and structure of data, provides the tool for fluent integration and use of data.

As discussed in section 2, data has meaning, a form and it can have a structure. When data is used in computing, it has to be defined so that it is usable for the intended purpose. When the use of data is relatively simple, e.g. in a spreadsheet document, it does not need to be specifically defined, but we use, e.g., data as a type of *currency* for calculating budgeting for a project, or the software we use automatically detects the type and sets it to the default currency. But even in this simple case, the data format is defined. When more data in complex form is to be used, e.g. new software applications are designed, data between different software systems are exchanged or a new database is designed, the definition or planning of data types and form is necessary. For this, information about the use of and the requirements for data is needed.

3.1.1 ISO 10303 STEP and data modelling with EXPRESS

In product or system design, a large amount of data is needed to describe all the design aspects, such as the geometry, assembly, instrumentation, or function and behaviour of the target. Typically, numerous different design and engineering software applications are used in the process and they have their native data formats. In addition, e.g. authorities can require that a certain type of design data is available throughout the product or system life cycle for any possible later needs. This raises the challenge of data interoperability and applicability. To ease the challenge, a large standardisation effort has been conducted, and it continues, to build a foundation of technologies and definitions, and the data formats for product life cycle data management. This effort is the STEP standard, or ISO 10303 *Industrial automation systems and integration – Product data representation and exchange* family of standards. The STEP standard provides means to model the data the EXPRESS data modelling language as well as dedicated data models for many application areas. The STEP standard is divided into several hundreds of parts, which all are separate standards. The overall structure of the STEP standard is described in standard ISO 10303-1 *Industrial automation systems and integration – Product data representation and exchange – Part 1: Overview and fundamental principles* (ISO 10303-1, 1994). The numbering of the parts of the standard follows the structure (ISO/TC 184/SC 4, 2004):

- Parts 11 to 14 specify the description methods,
- Parts 21 to 28 specify the implementation methods,
- Parts 31 to 35 specify the conformance testing methodology and framework,
- Parts 41 to 58 specify the integrated generic resources,
- Parts 101 to 110 specify the integrated application resources,
- Parts 201 to 240 specify the application protocols,

- Parts 301 to 336 specify the abstract test suites,
- Parts 501 to 523 specify the application interpreted constructs, and
- Parts 1001 to 1514 specify the application modules.

The EXPRESS data modelling language is defined in standard ISO 10303-11 *Industrial automation systems and integration -- Product data representation and exchange -- Part 11: Description methods*: The EXPRESS language reference manual (ISO 10303-11, 2004). The EXPRESS language is used for defining the data models for all the STEP family standards. The EXPRESS modelling language has also a graphical notation for graphs, EXPRESS-G, and XML representation of EXPRESS schemas and data (ISO 10303.28, 2007).

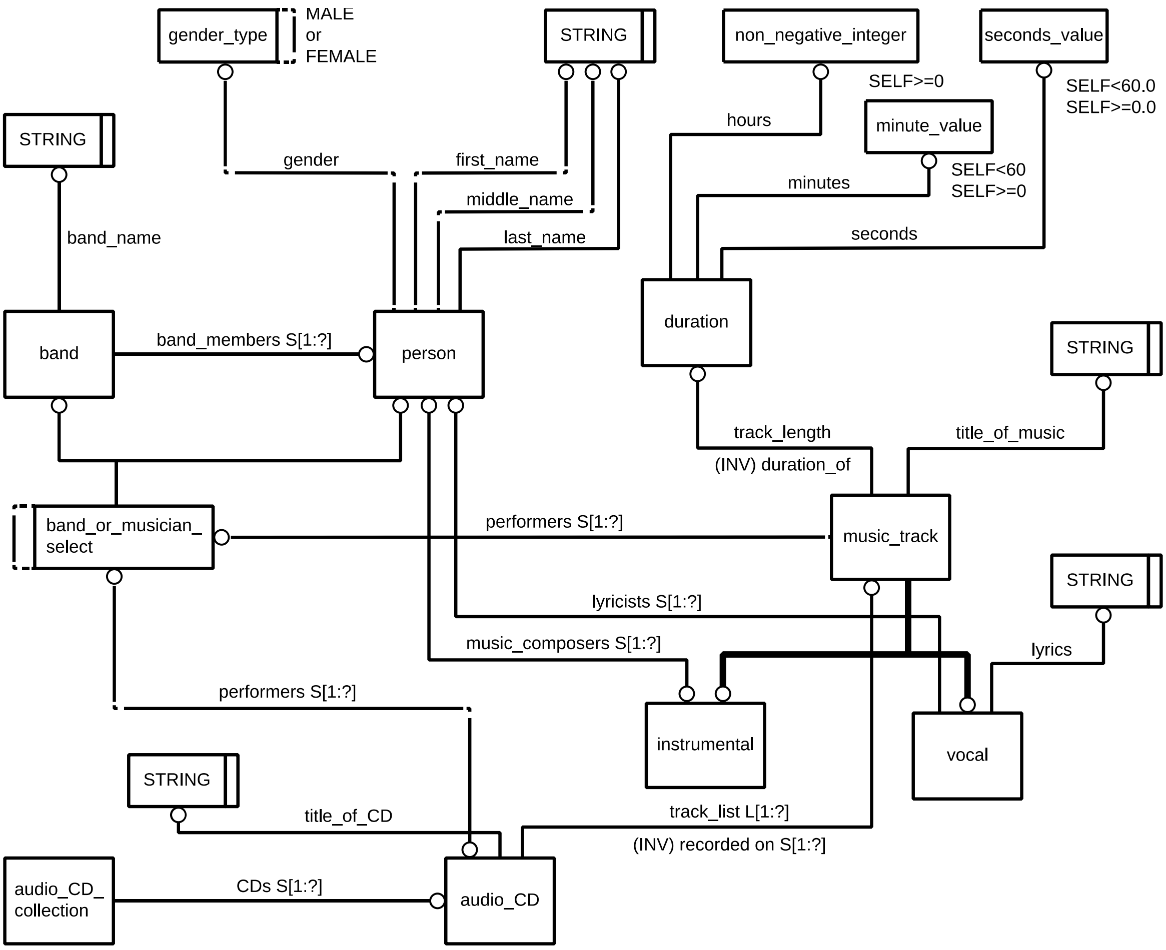


Figure 9: An example of a graphical data model in EXPRESS-G notation (source: https://en.wikipedia.org/wiki/EXPRESS_(data_modelling_language)).

In general, a modelling language for data modelling provides additional tools for software development, such as data validation and automated processes to define the data structures in software applications.

3.1.2 The Semantic Web

Data modelling is a fundamental part for semantic data representation in the Semantic Web, a semantic layer of the current Internet. The Semantic Web aims at having mechanisms for representing the meaning of data together with data and providing means to query and reason the data to better match the needs of the user. The Semantic Web defines a set of technologies for data, information and knowledge representation, for queries, for setting rules on the data and for reasoning. The base technology for the Semantic Web is the Resource Description

Framework (RDF) (World Wide Web Consortium, 2014). RDF specification defines the fundamental concepts for data representation, of which the subject-predicate-object model is the most important (Figure *10*). The model defines how the data elements are linked and how the data components, i.e. objects and their relations, are defined and used.

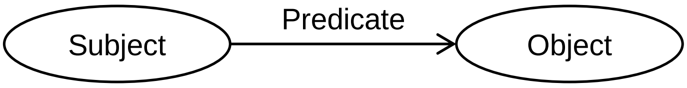


Figure 10: The subject-predicate-object model of the Resource Description Framework (RDF).

The data models, i.e. ontologies, in the Semantic Web context, are defined with the Web Ontology Language, OWL (World Wide Web Consortium, 2012). The Web Ontology Language defines the syntaxes, semantics and structures to be used for defining the data modelling components in the ontology. An ontology can be seen as the definition of the data modelling elements for the given content. An example of an ontology is simply contact information, organisation and project ontology that contains the following objects types and their attributes:

- **“person”** with attributes:
    - “first_name”
    - “surname”
    - “phone_number”
    - “email_address”
    - “position”

- “organisation_unit” with attributes:
    - “unit_type”
    - “unit_name”
    - “unit_code”

- **“project”** with attributes
    - “purpose”
    - “start_date”
    - “end_date”
    - “volume”

and relations:

- “has-a-unit”
- “belongs-to-unit”
- “is-a-supervisor-of”
- “has-a-project”
- “is-a-member-of”
- “is-a-project-manager-of”

In this example, we have only three object types and six relations, which can be used for modelling the data to describe an organisation together with the contact information of the people in it and the projects they are working in. Even this small set of modelling elements, i.e. the object and relation types in the ontology, enables modelling of very complex data models. An example of a data model modelled with the above described elements is shown in Figure *11*, Figure *12* and Figure *13*. The figures do not contain the data properties, i.e. the actual data content, of the modelling elements but only the main data model elements and their relations.

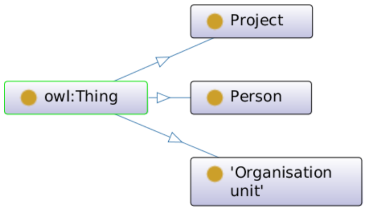


Figure 11: A visualisation of the contact information, organisation and project ontology.

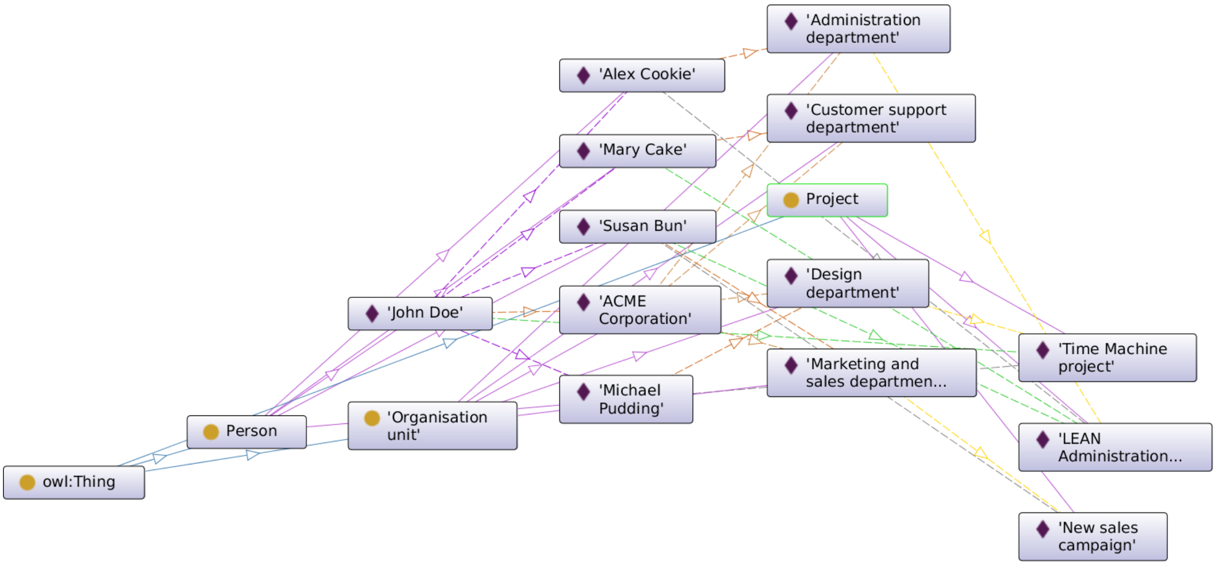


Figure 12: A visualisation of the contact information, organisation and project ontology, and a data model utilising the ontology, i.e. the individuals (instances) of the object and relation types.

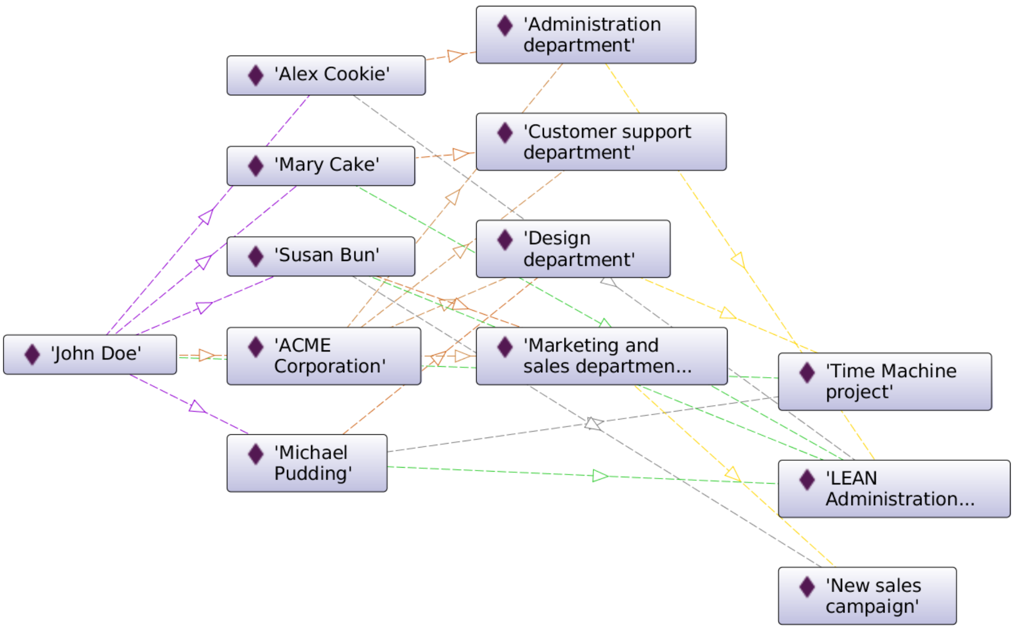


Figure 13: A visualisation of the individuals (i.e. instances) of a data model utilising the contact information, organisation and project ontology.

There are multiple strategies available for transforming data into ontology. For example, Entity Relationship Diagrams (ERD), which show the relationships of entity sets stored in a database, can be transformed into ontology graphs (Fahad, 2008). In addition, formal concept analysis methods can be used to map contents in the database to the ontology. In the formal concept analysis, concept hierarchies are extracted from datasets using mathematical models (Gao et al., 2016). Finally, XML can be used as a means to both of these approaches to transform the data into any other computer platform-based data form (Malik et al., 2016).

The Extensible Stylesheet Language family (XSL, a set of W3C specifications, https://www.w3.org/Style/XSL/) defines specifications, technologies and data models for data transformations. The main target is data transformation from one data model in XML format to another data model either in XML or other format. In principle, the approach enables data transformation from any format to any other format.

Linked Open Data (LOD) is an open, interlinked collection of datasets in machine-interpretable form, covering multiple domains from life sciences to government data (Rosati et al., 2016). At present, there exists several tools and frameworks for mapping relational databases to LOD. For example, the D2RQ tool defines a declarative language to describe mappings between application-specific relational database schemata and RDF-S/OWL ontologies (Ristoski & Paulheim, 2016). In general, LOD-based approaches can be utilised to interpret different types of data including structured, semi-structured and unstructured data (Ristoski & Paulheim, 2016).

3.2 Data exchange, transformation and integration

Data exchange means, in general, the action of exchanging data between software applications or data management systems. Data exchange is needed, e.g., when some software application is used for modifying data, which is then used in other software applications or systems. Data integration means using the data managed in one software application or a system in other software applications or systems. An example of data integration is to use the data, managed in an enterprise resource planning (ERP) system, in other data management systems, such as in a project database. Data transformation is the process of converting data from one format or structure into another format or structure. Data transformation is critical to activities, such as data integration and data management. Data transformation can become necessary in many situations but most typically, transformations are needed when data is made compatible with other data, e.g., a legacy system is migrated to a new information system, or multiple data sources are to be integrated (Rahm & Do, 2000).

In design and engineering, data integration and thus transformation needs occur when design and engineering tools and systems are used in a chain, i.e. the output of one tool or data stored in one system is used as the input or is stored to another system. Examples of engineering tools are computer-aided engineering (CAE) tools, such as structural analysis and system simulation software applications. Examples of engineering and design systems are product life cycle management (PLM) and simulation life cycle management (SLM) systems. The need to increasingly utilise computational approaches in design and engineering, and to automate common design and engineering tasks increase the need for fluent data exchange and data integration.

Today, many companies are increasingly utilising a variety of services provided by the biggest and most-influential cloud-computing vendors including Microsoft, Amazon and IBM, for example. While previously organisations had to buy and maintain their own physical servers, cloud computing enables companies to consume computing resources as a utility – just like electricity – rather than having to build and maintain computing infrastructures in-house (Ochs & Riemann, 2018). Cloud computing is emerging as one of the major enablers also in the manufacturing industry; it can replace traditional manufacturing business models with more agile, scalable and efficient business practices, help it to align product innovation with business strategy, and create intelligent factory networks that encourage effective collaboration (Xu, 2012).

In cloud computing, the heterogeneity and lack of standardisation hampers the use of cloud services offered by multiple providers (Cavalcante et al., 2016). Therefore, the importance of cloud interoperability has been highlighted by both the industry and academia and many cloud interoperability standards have been proposed (Zhang et al., 2013). However, it may take years for the standards to be fully agreed upon and adopted, if ever, and as practical, short-term solutions, technologies have been developed to enable interoperation among clouds, from both the cloud provider's and user's perspectives (Zhang et al., 2013). Data transformation is seen as one of the key technologies that can facilitate the migration and integration of applications and data between different cloud service providers (Carrasco et al., 2014).

The naive way to translate data from one format to another is writing a specific program for each translation task and writing such a program is typically a non-trivial task, often complicated by numerous technical aspects of the specific data sources that are not relevant to the translation process (Abiteboul et al., 1999). A sound solution for a data integration tasks requires a clean abstraction of the different formats in which data are stored, and means for specifying the correspondences between data in different contexts and for translating data from one context to another (Abiteboul et al., 1999).

Data transformation can be examined from multiple viewpoints. For example, there are fundamental differences between processing statically stored data and streamed data arriving from IoT devices or sensors. Moreover, data transformation can be applied for statistical analysis (e.g. measurement variables do not fit a normal distribution or have greatly different standard deviations in different groups) or for improving interoperability, (e.g. data sets are heterogeneous in terms of vocabularies, formats, data representation and modality). In this document, the focus is on the latter perspective.

The constant increase in the volume of data caused by, for example, the emergence of social media, IoT, and multimedia, has delivered an overwhelming stream of data in either organised or unstructured formats. In general, devices providing streamed data are extremely heterogeneous regarding basic communication protocols, data formats, and technologies and, in addition, the systems involve diverse modalities (Cheng et al., 2019). For example, the current smart phones are equipped with several different sensors, such as an accelerometer, digital compass, gyroscope, GPS, microphone and camera (Lane et al., 2010). Moreover, the IoT data can be collected from many different sources and consist of various structured and unstructured data (Jiang et al., 2014). Data transformation and harmonisation can facilitate data interoperability and allow sharing and use of data across multiple services.

The statically stored data represented, for example, in relational databases is usually backed by a schema, which formally defines the entities and relations between them. In most of the cases, the schema is specific for each database, which does not allow for automatic data integration from multiple databases (Datta et al., 2016). For easier and automatic data integration and extension, a global shared schema definition should be used across databases (Ristoski and Paulheim, 2016). In order to comprehend data and to extract knowledge, the data must be sorted, changed, blended and prepared both measurably and logically (Malik et al., 2016). In the following paragraphs, some methods and techniques for data transformation are briefly discussed.

One of the most widely applied methods for data transformation is the utilisation of semantic technologies, and in particular, ontologies. A commonly agreed definition of ontology, made by Gruber (1993), is the following: "An ontology is an explicit and formal specification of a conceptualisation of a domain of interest". Furthermore, ontology is defined as a controlled vocabulary that describes objects and the relations between them in a formal way; ontology resembles faceted taxonomy but uses richer semantic relationships between terms and attributes, as well as strict rules about how to specify terms and relationships (Uschold and Gruninger, 1996; Berners-Lee et al., 2001). Semantic data and ontologies may contain information about data transformation, or data mapping, e.g. by providing information about equivalence of concepts and relations. In addition, separate ontologies can define equivalences between concepts and relations of two or more other ontologies. There are also software tools and services for semantic data mapping and transformation.

3.3 Cleaning up data

Raw data is rarely useful as such and might require a lot of work before intended data analysis can be conducted. Quality of the dataset is crucial for the possibility of extracting information from the data. Thus, content of the dataset must be critically assessed because several issues can deteriorate the quality of data. Reasons can be random or systematic. Random errors, if within a reasonable amount, are not typically crucial for data analysis because those do not cause bias in the results. Random errors can be traced, for example, by visualisations or cross-tabulations.

Systematic errors can cause biased results and to avoid misleading results, the collected dataset must be studied carefully. The main reason for systematic errors can be found from data collection either from official instructions or recording practices. To be able to recognise systematic errors, the analyser needs a deep understanding of how data have been collected, which information variables really contain, who made the entries, when were those made, what are the recording practises, etc. It is important to determine data collection practises not only to study official guidelines because practices are not always in line with given instructions (Kortelainen et al., 2015).

3.4 Creating value from data

Data is the fuel of any intelligent industrial system. Data can be seen as a raw material which companies collect, acquire or generate. This data must then be refined further into information or knowledge that holds more value for the data user or owner. This process is called the data value creation process.

3.4.1 Data value creation process

The value creation process typically starts with identifying relevant data sources, continues to the generation of knowledge by refining, analysing and modelling of the data and leads to an output that creates value for the products, services and operations of the company. An example of such process is presented in Figure 14. In addition, the process should incorporate managing knowledge and skills necessary for the analysis.

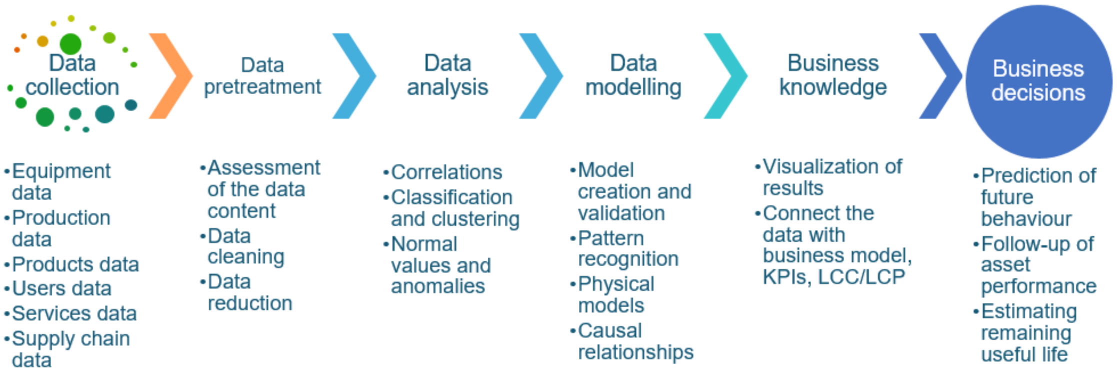


Figure 14: Data value creation process (modified from Kortelainen et al., 2017).

Whereas Figure 14 highlights the business perspective to the value creation process, the IT view is different. From the IT perspective, the process includes extraction, transformation and loading of data into a database, analytics software, user interfaces, proactive utilisation of analytics results by forwarding them to operational systems, and automating knowledge and skills related to the data analysis. Big data management canvas (Kaufman, 2019) is an approach for illustrating data refinement and value creation that connects business with the information technology perspective. In this approach, the steps including data preparation, analysis, interaction, effectuation, and intelligence align both business and IT to create value from data.

3.4.2 Data exploration and descriptive data analysis

Data exploration aims to determine and understand the nature and characteristics of data. Descriptive data analysis pursues revealing the information that the collected data contains and develop insights for the user. An example is presented in Figure *15* that visualises increasing the failure rate of an ageing system. Statistical data analysis methods can be used to compare data groups, find correlations or other relationships between variables, define clusters, find normal values and anomalies, detect trends or other patterns and so on.

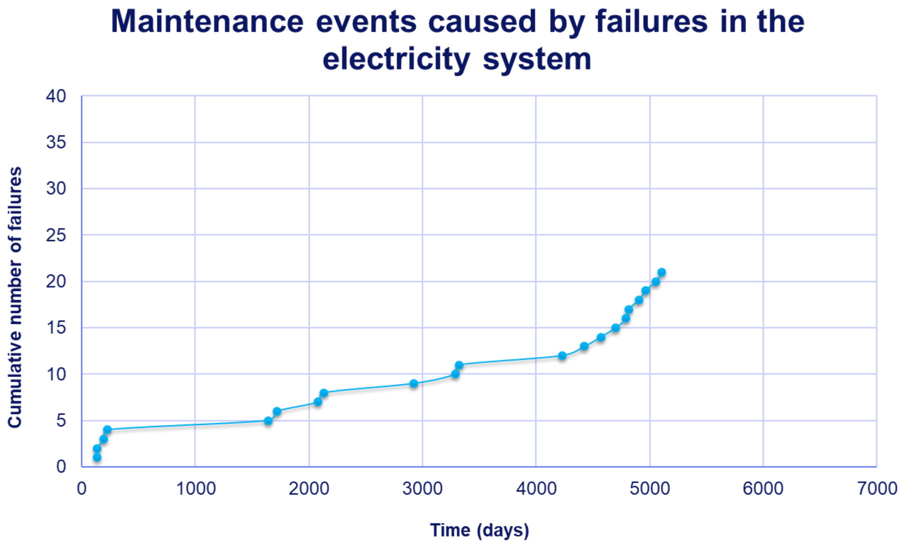


Figure 15: An example of a descriptive analysis of maintenance data.

Descriptive data analysis provides summaries of collected data. Summaries can be basic measures (e.g. mean, median, variance) describing data, tables (e.g. frequencies, cross-tabulations) or figures (e.g. bar charts, box-plots, spider/radar charts) presenting relevant features of the data. Descriptive data analysis creates understanding of the data content and insights into interesting relationships between variables.

3.4.3 Predictive data modelling

Data modelling requires different capabilities, concepts, methods and methodologies, algorithms, models and tools depending on the scope of the analysis. Various data analysis methods exist and a suitable method for a case is dependent on the problem and available data. This chapter does not provide a review of existing data analysis methods, but raises the importance of understanding the behaviour of a target system and how qualitative analysis can support statistical data analysis.

The real added value of data is in exploiting it to predict future behaviour and follow the performance of the assets, to estimate remaining useful life, to identify the cause of underperforming systems and to support planning and decision-making. Figure *16* presents an example of the data modelling in two scenarios: failure rate without any actions and failure rate after system replacement. Complex systems require complex models, as they consist of several different components whose specific behaviours and interactions needs to be understood and modelled. In order to translate the patterns, anomalies and trends to predictions of remaining lifetime or future behaviour of the item, further information about the system is needed.

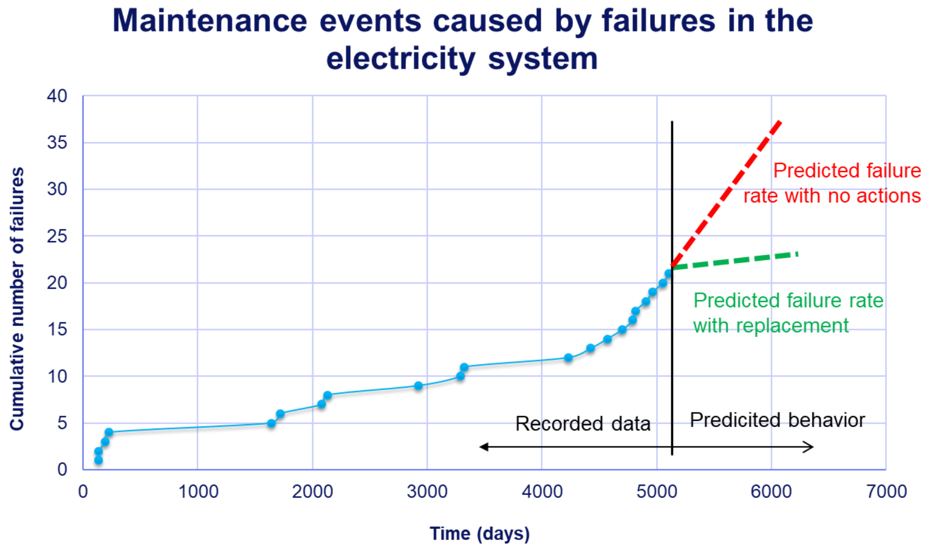


Figure 16: Examples of data modelling and predicting future behaviour with two scenarios.

Traditional physical models are highly complicated and require a lot of modelling efforts, or models are too simplified making it impossible to capture relevant behaviour. Data-driven models and prognostics and health management (PHM) algorithms usually use pattern recognition and machine learning techniques to detect changes in system states. Also, qualitative information like risk and reliability analyses, e.g. Hazop and FMECA, support the analysis phase providing essential information about the target application. These analyses could provide cause-consequence chains that connect failure indication or initiation pattern or a deviation from a certain chain of events, and link the emerging event with expected consequences. This allows the user to make predictions and to take proactive actions in time.

3.4.4 Supporting business decisions

A further step to add value is to connect the data with business-related information like a business model, key performance indicator framework, life-cycle cost and profit model, or decision-making situation. The history data alone is not a sufficient basis for future forecasts but tacit knowledge, expert judgment and other data sources that help to understand customer´s value creation have to be utilised. The ‘data analysis’ may then take a variety of forms and actions from benchmarking and traffic light dashboards to calculations that support long-term decisions like investments.

The human brain typically is not capable of finding out relevant information from a set of numbers which is in a general form for collected data. Visualisations are one tool to present data in an understandable way. The main aim of the visualisations is to condense a large amount of data in a form which supports decision-makers’ or other users’ attempts to understand the information included in the data. Visualisations are important tools for communicating findings between data analysists and decision-makers.

4. Industrial Internet of Things

Much of the change in design and engineering, manufacturing and services is driven by the technological development regarding the access, communication and analysis of large amounts of data. The different systems that ideally make design and engineering, manufacturing and service operations more effective and efficient are evolving rapidly, becoming more powerful and complex. With this development, a new challenge emerges at the intersection of the different systems. In the new smart design and engineering, manufacturing and service management systems, several components have to be integrated and work together as seamlessly as possible. This comes down to the Integration of Information Technology (IT), Operational Technology (OT) and Engineering Technology (ET) (World Manufacturing Forum, 2018).

IT/OT/ET convergence is the integration of information technology (IT) systems used for data-centric computing with operational technology (OT) systems used to monitor events, processes and devices, and make adjustments in enterprise and industrial operations and engineering technologies (ET) with Digital Engineering Models, providing various analytics and simulation capabilities seen in Figure 17.

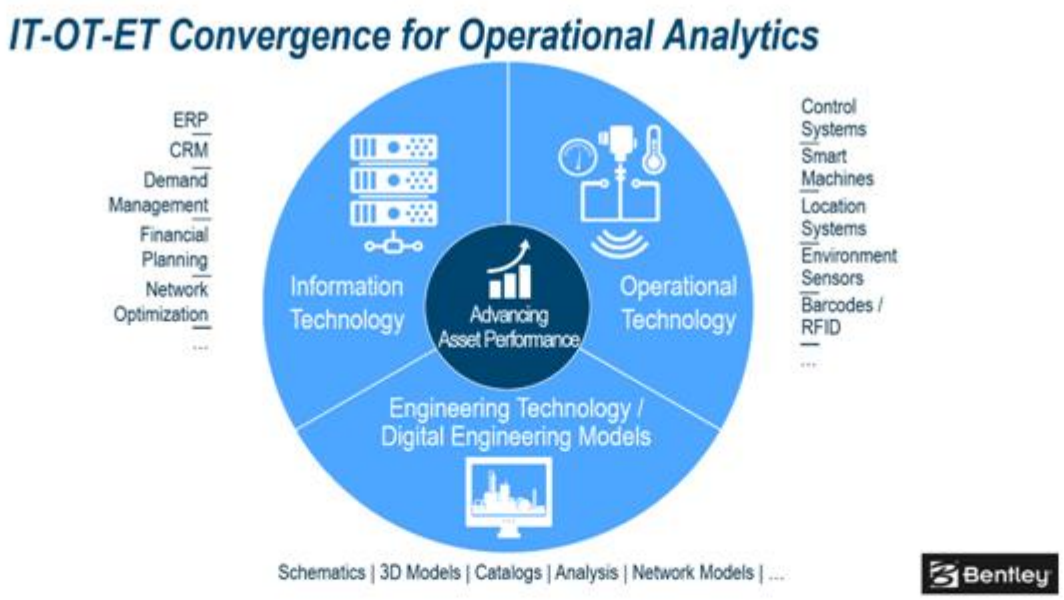


Figure 17. Convergence of information technology (IT) and operational technology (OT) systems with engineering technologies (ET) for operational analytics (Singh, 2017).

In the manufacturing industry and related processes, an increasing amount of data is generated by physical assets. Intelligent, networked products and services require companies to build sensors embedded in physical assets. Intelligent assets require a completely new, multi-layered technology infrastructure (technology stack, Figure 18). The infrastructure consists of software, applications, networks, equipment, product cloud, information management platforms, and business processes and processes based on them.

Industrial internet of things (IIoT) can be described as a stack shown in Figure 18. At the base of the stack there is the physical world consisting of tangible things such as machines. They interface the digital world via sensors, which measure the phenomena of the real world and actuators impacting the real world. The function of the next layer of the stack is to transmit data from sensors to the IoT (IT / OT) platform and vice versa. The stack also supports Industry 4.0 that is necessary for the smart manufacturing (see Figure 25).

The function of the IoT platform and data-analytics layers is to manage, store and analyse the data. In practise, these layers are implemented with commercial IoT data platforms. These layers are the most relevant for this report. The operative layer may be implemented as human or automatic decision-making.

The Industrial Internet Consortium (www.iiconsortium.org) has published an Industrial Internet Reference Architecture (IIRA, v1.9) (https://www.iiconsortium.org/IIRA.htm) which describes the issue from multiple viewpoints.

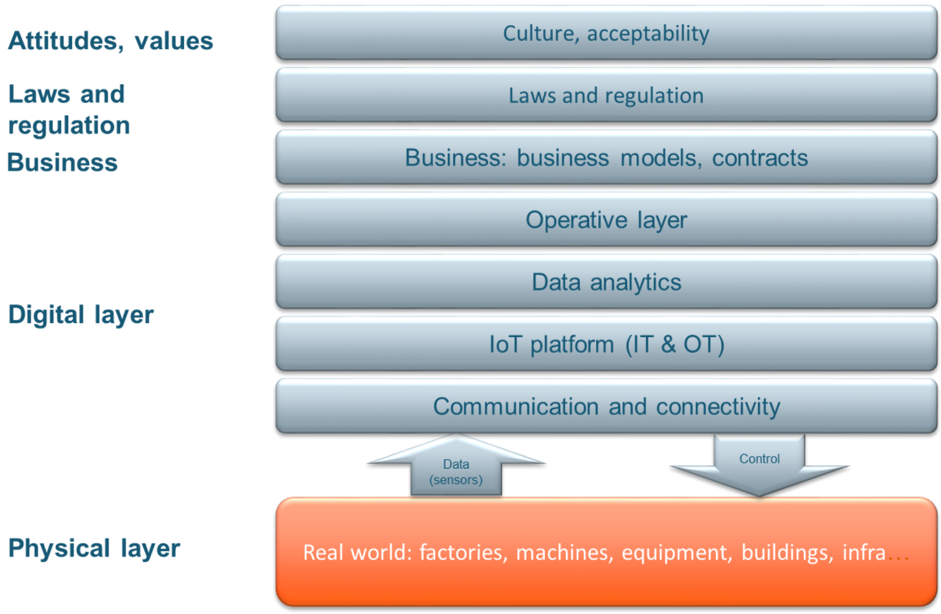


Figure 18. Technology stack including cultural, social and regulatory layers.

5. Artificial intelligence and data analytics

The increasing use of Artificial Intelligence (AI) technologies will further transform the ways in which data is utilised in industry. The concept of AI, however, does not have a single agreed definition. Often, AI is defined broadly as a system capable of understanding its environment and making rational decisions accordingly (e.g. Nilsson, 2011). In this context, we focus especially on data-intensive AI approaches, and especially ones applying machine learning methods to vast amounts of industrial data.

In recent years, the hype around AI has been largely based on advances in machine learning. Especially results in fields like image recognition and natural language processing have kept AI constantly in the headlines. This is partly enabled by advances in the actual machine learning methods, but the rapid development is also enabled by the increasing amount of available data, as well as improvements in data processing capacity and its declining cost. It should be noted, however, that various model-based and rule-based systems will remain to be relevant also in the future, despite the imminent increase of machine learning technologies. In many industrial applications, hybrid approaches combining data-based and model-based approaches are likely to be beneficial.

The concept of machine learning can further be divided into the following fundamental paradigms (Jordan & Mitchell, 2015), which all have potential uses in the manufacturing industry:

- **Supervised learning** is used when a machine learning model can be taught using an example set of desired outputs, such as a set of annotated photographs when teaching an image recognition system. Typical uses of supervised learning include classification and regression problems. Supervised learning often applies artificial neural networks, which loosely mimic the way neurons work in human brains. Applications using multi-layered deep neural networks with large amounts of data are also referred to as deep learning (Lee et al., 2018). In the manufacturing industry, supervised learning is currently the predominant paradigm, as it has many practical uses, for example in quality control, predictive maintenance and process optimisation tasks (Wuest, 2016).
- **Unsupervised learning** is used when unstructured data is available (i.e. no datasets are available or practicable for teaching the system in advance). Thus, the approach is

especially useful in exploring the data, e.g., to find clusters and make categorisations, or to detect anomalies in the data (Lee et al., 2018).

- **Reinforcement learning**, sometimes also categorised as a subset of supervised learning, is based on the interaction of the AI system and its environment (Antonoglou et al., 2015). It is especially interesting from the point of view of future autonomous technologies and collaborative robotics, as the approach enables continuous learning during operations.

The power of AI lies in its applications to problems where it outperforms humans. In recent years, the equivalence with human performance has already been reached in several limited tasks. These tasks, such as specific image or speech recognition tasks, are rather narrow and have a very defined scope. In more general problem-solving tasks, humans are expected to remain far superior to AI systems in the foreseeable future, but they can still benefit greatly from decision-making support provided by AI and data analytics.

The use of AI in the manufacturing industry has been studied in several instances, especially as a part of Industry 4.0 considerations in an environment where IoT, connectivity and big data collection work in collaboration with AI systems (Li, 2017). Some expected advantages include (adopted from Wuest (2016), Richter & Schmitz (2019)):

- Improved decision-making support capability when compared to traditional systems, for example in predictive maintenance and lifetime prediction, supply chain management, and quality control,

- Decreasing reliance on experienced operators and decreasing impact of the variations in operators’ qualifications in terms of quality and productivity,

- Quick adjustments in manufacturing strategy and production plans,

- Potential occupational safety benefits as humans can be removed from dangerous environments.

Despite the great promises for increased productivity and safety, there are still unsolved issues in applications of AI in industry. The major challenges include both technological as well as wider socio-technical challenges. For example, several concrete safety challenges (see e.g. Amodei et al., 2016) are related to AI systems and their application in industrial systems. Transparency and intuitive ways for human-AI interaction are crucial factors in enabling trust for AI systems (Karvonen et al., 2019). Testing and validation of AI systems will also require new approaches. Furthermore, the increasing application of AI sparks several wider questions regarding the societal impact and ethicality of the use of AI (Leikas et al., 2018). In the manufacturing industry, this can be seen also as a continuation of the discussion around the effects of automatisation and robotisation.

The introduction of AI into manufacturing processes is likely to be gradual, starting from various optimisation and decision-making support tasks, and advancing towards increasingly autonomous and AI-operated systems. Machine learning relies on data, meaning that the availability and quality of data, as discussed in this report, become even more important.

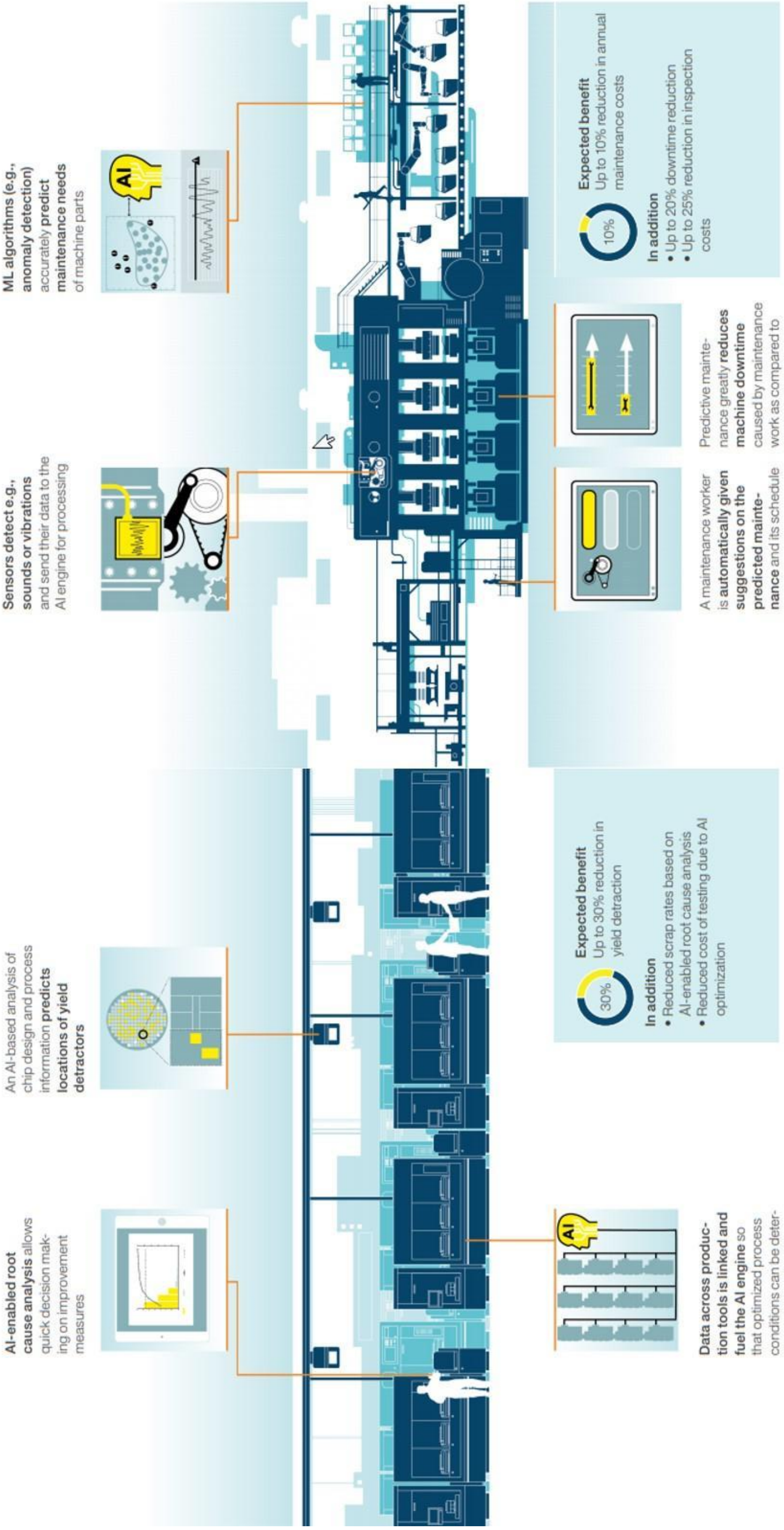


Figure 19. Examples of using AI in manufacturing industry (McKinsey, 2017).

5.1.1 Data transformation using AI

In addition to implementing AI in the actual industrial processes, it can also be applied as a tool for data transformation to support other data-related activities. For example, several approaches exist that utilise ontologies and artificial intelligence for data harmonisation (Ping et al., 2018; Bruseker et al., 2017). On the other hand, computation models for the fusion of multimodal sensory data have been constructed according to the hidden Markov Process (Cheng et al., 2019). Moreover, Amin et al. (2018) introduced ‘Rank’ and ‘Log’ data translation methods to improve the performance of cross-company prediction models when one company’s (source) data is used as a training set and another company’s (target) data is considered for testing purpose.

6. Challenges and opportunities in using data

The manufacturing industry is a main focus in this paper, covering the whole product or system life cycle (from the pre-concept design phase, through design & engineering, manufacturing/production, use/operation, together with maintenance and support services, until the end-of-life of the product or system). The life cycle phases are selected for the following examples: 1) engineering & design, 2) manufacturing 3) supply chain management and 4) asset management and life cycle services. The connecting elements over life cycle phases are the product/system and the related data. The following chapters highlight the challenges and opportunities related to using data and to those of creating and refining the data.

6.1 Design and engineering

In a typical computational engineering process, several computational methods are used for modelling, analysing and simulating the function, behaviour and dynamics of the target under design. The methods may be, e.g., the finite element method (FEM) for structural analysis and analysis of structural dynamics, computational fluid dynamics (CFD) for fluid flow simulation, multibody system (MBS) simulation for the simulation and analysis of machine system dynamics. In most of these methods, 3D data of the product or system components is needed, and the source of the data is the CAD tool or the PLM system. In addition, the output of one analysis or simulation tool may be used as the input of another. For example, the flexibility of a machine part is analysed with a FEM software application and the data is then exported to an MBS software application, where it is used as a component of a simulation model to simulate the dynamics of the system. Furthermore, the output of an MBS simulation can be used as an input for a CFD simulation of the system. There are numerous software applications available for each computational method and they all have their own dedicated native data representation. Standardisation is available for only some of the computational domains.

In the technology industry, using subcontracting is a common practice, especially when rarely needed demanding engineering analyses are required. This complicates the data exchange and interoperability of tools and systems even further, as different parties are more likely to have different tools for the same computing purpose.

6.2 Manufacturing

Tao et al. (2018) divided the meaning and the use of manufacturing data into four ages: the handicraft, the machine, the information and the big data age (Figure *20*). While craft production relied on undocumented experience and tacit knowledge, the modern sales-delivery channel is loaded with the high volume of data. The configured product definition data must be handed over to the upstream manufacturing operations. Consequently, manufacturing the instances of a product family delivers a plethora of associated data. On top of increased volume, there is a proliferation of the breadth of manufacturing information, i.e. the types of data. Along with this, the granularity of manufacturing information has increased. Archives, documents and binders have been replaced by datasets in databases, and nowadays the instances of data are stored in a cloud or on the internet.

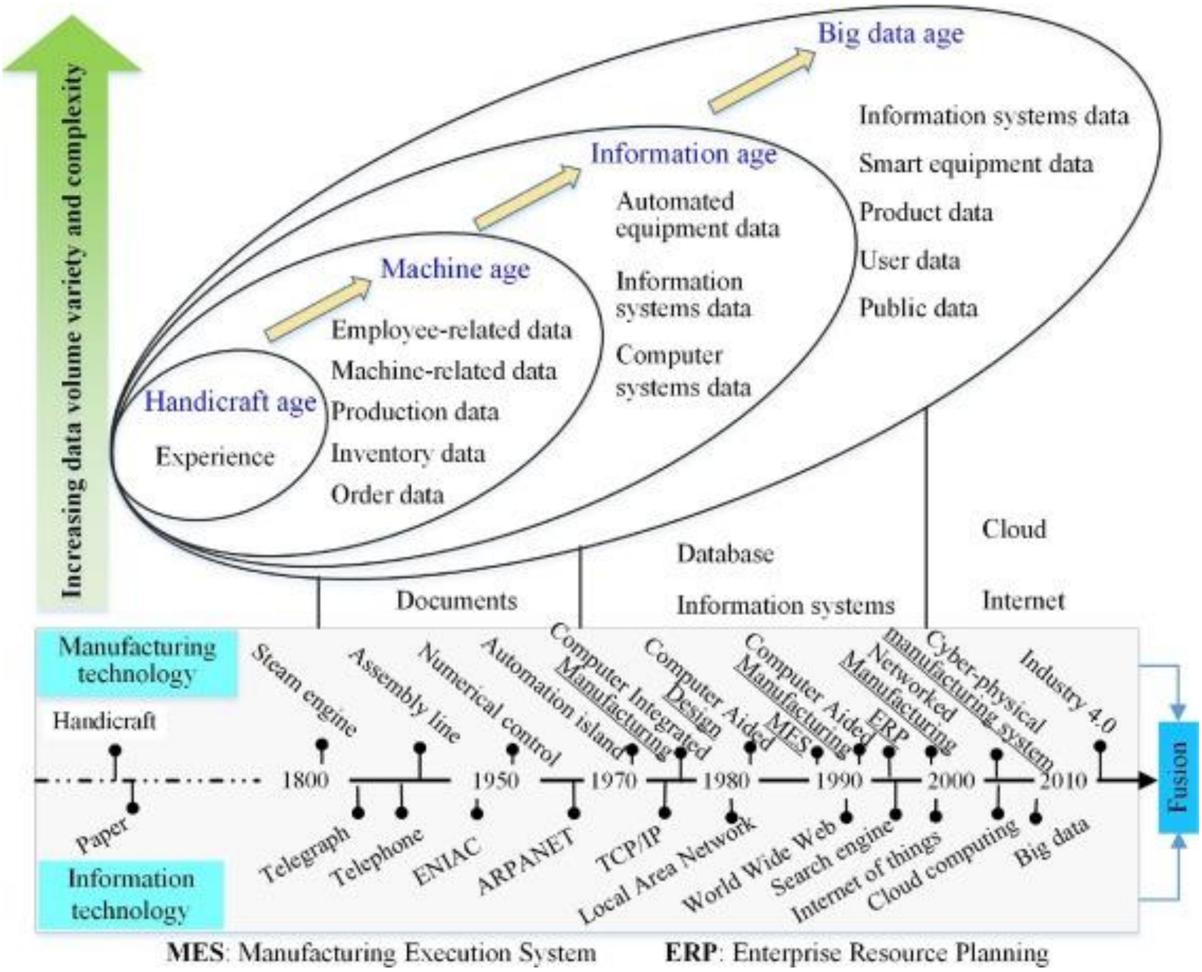


Figure 20. Evolution of data in manufacturing (F. Tao et al., 2018, pp. 159)

Throughout the ages, agility has been a persistent imperative for industrial operations. The need to be agile was recognised before craft production and it existed throughout industrialisation and mass production. Being agile and effective was one of the key drivers for Toyota Production Systems (TPS), which is known in the Western literature as Lean Manufacturing (Olesen, 1998). For example, the increased agility of production appeared in the minimal changeover time of dies that was one of the innovations by Ohno (Klemke & Nyhuis, 2009) or SMED (Single Minute Exchange of Die) by Shingo (1989).

The limited, average response to the variety of customer needs was a trade-off for efficiency in Mass Production. Customisation was neither the key driver for TPS or Lean, but rather a response to the effect of increased product variation that was cast upon manufacturing. However, mass customisation of products with the high variety of products is an answer to customer-oriented agility. Customisation can be considered as the genuine reason for manufacturing agility. There, the lot size of 1 is the requirement for the final assembly.

6.2.1 Product Lifecycle Management: definitions, structures, variety and closed-loop

Product life cycle management (PLM) "includes all organisational tasks necessary for the identification, supply and archival storage of product-related data during the product life cycle" (Hirz et al., 2013, p. 43). It is often related to PLM systems that actually often include integration into different engineering software, such as computer aided design, engineering and manufacturing applications (CAD, CAE, CAM). However, PLM is a larger concept than just software, because the decisions on PLM strategy and PLM processes have wide effects on the product and manufacturing strategy and also on the actual engineering and production processes.

As stated in chapter 1.3.2 the product definition is a key element for manufacturing operations. In the PLM systems, product definition datasets are related to the items of the Bill of Materials in the forms of models and drawings. Even today, the electronic counterpart, a kind of digital twin of paper-based documentation is dominant in manufacturing, but a more compact form of product definition is emerging (see Figure *21*).

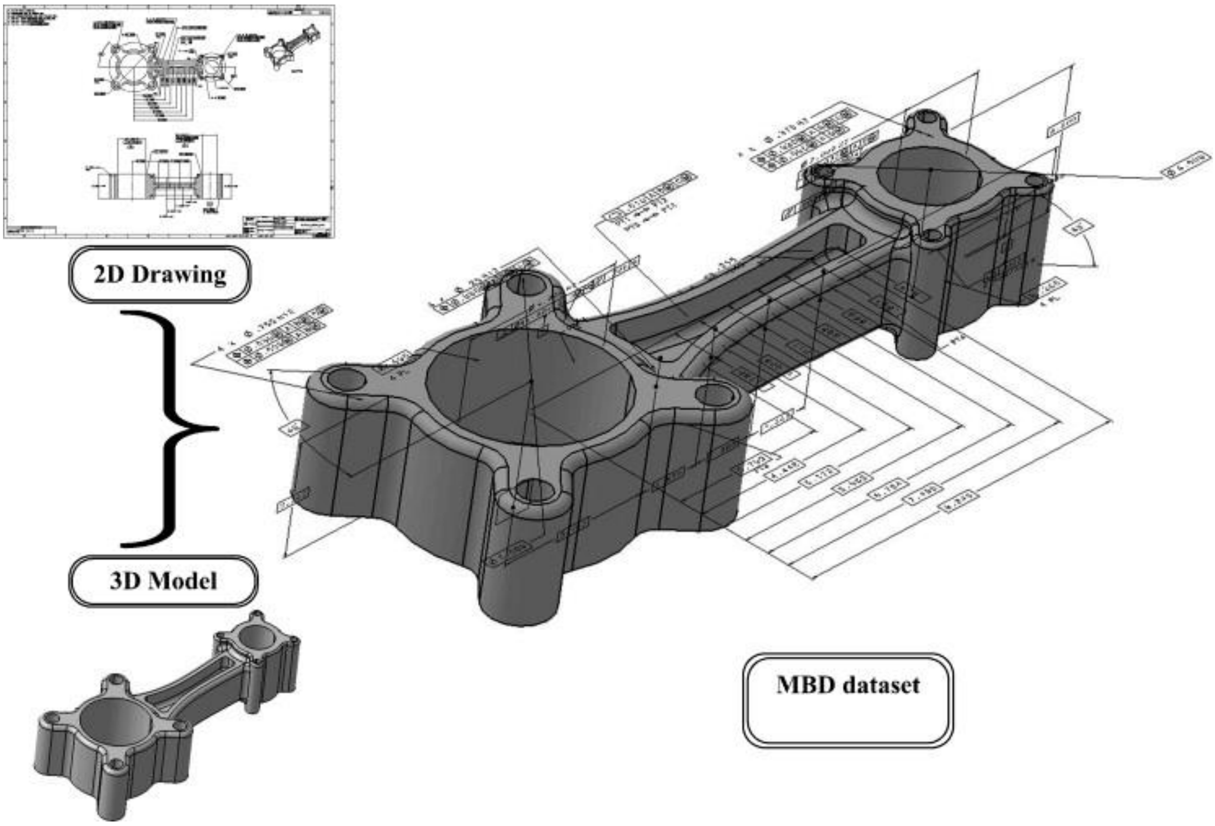


Figure 21. Example of an MBD dataset (by Quintana et al., 2010, p. 499 https://doi.org/10.1016/j.compind.2010.01.005).

A Model-Based Definition can serve as an integrated presentation of a part that replaces the dataset of a separate model and drawing. Thus, it enhances the idea of a single source of truth, which is one of the overall ideas behind PLM. Companies expect significant changes in the way engineers carry out their work when MBD is adopted (see Figure *7*, p. 14).

For manufacturing operations and production planning, such as computer aided manufacturing, it is favourable that an MBD dataset is in machine-readable form, a representation of product definition. This approach can ensure the seamless single source of product definition and minimise unnecessary errors due to human intervention, e.g., in off-line programming of toolpaths. MBD datasets can be useful also in quality control as the geometry and associated attributes, such as dimensions, geometric tolerances, basic elements and surface qualities, are integrated into an MBD model.

As MBD primarily addresses part of manufacturing processes, mass customisation relies on proactive agility that is enabled by the single source of product variety definition that is built into PLM processes and support (see Figure *22*). Thus, the many capabilities related to the management of product families and variation within the definition of product families can be seen as one of the key enablers for the agility of product customisation. There, capturing the knowledge on product variety is necessary.

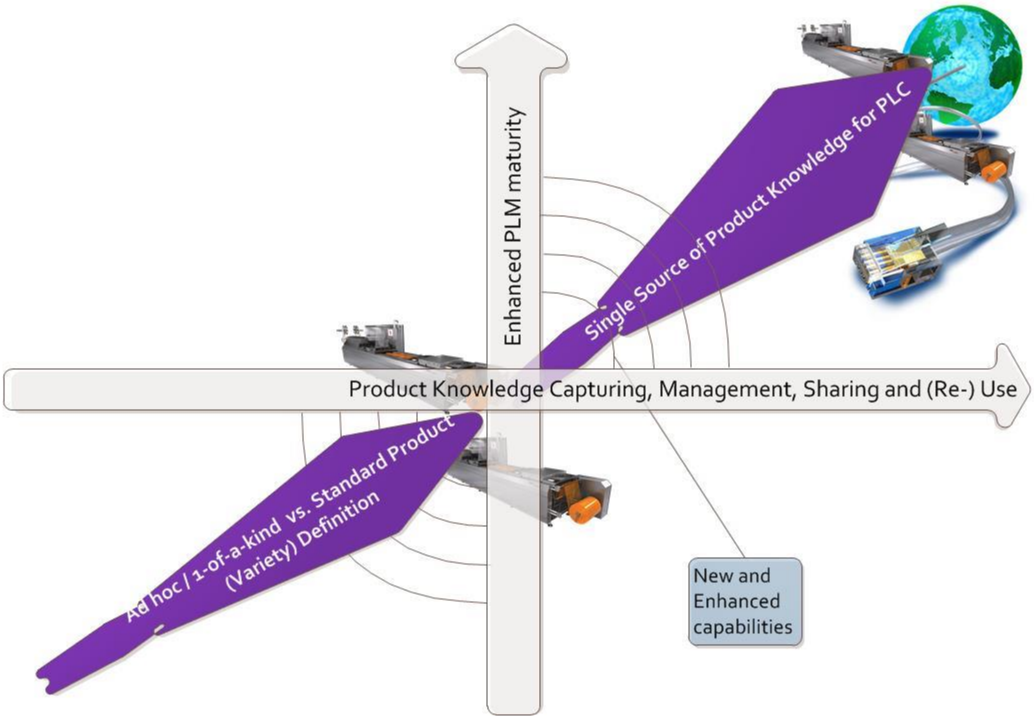


Figure 22. PLM maturity for enhancing the capabilities of capturing, management, sharing and re-using product definition (Pulkkinen et al., 2017).

The means for managing a product variety definition are product configuration systems or simply product configurators (Blecker et al., 2005). The primary concepts are objects in part-of and kind-of structures in different domains (Yu & MacCallum, 1996, Harlou, 2006) and different kinds of rules / constraints between the objects of different types. For the capturing of configuration knowledge, different sources suggest object-oriented analysis and modelling. The configuration ontology (by Soininen et al., 1998) consists of the structures and constraints of different types: ports, resources, contexts and generic constraints. The approach by Mortensen et al. (2008) utilises different modelling techniques and especially, for the capturing of configuration knowledge, the consistent structures of three domains (Harlou, 2006). His idea is to define re-usable design assets that are called standard designs. Eventually, a product configuration is an instance of a configurable product family, i.e., a definition of an individual instance of all possible instances.

The set of life-cycle processes require the use of dedicated data management systems. In an organisation this leads to a unique data management architecture with a set of limited functionally separate platforms, such as product life-cycle management, enterprise resource planning and manufacturing operations management systems. The challenge is not only the increased volume, variety and granularity of data, but also the different formats and versions of the product, business operations and manufacturing data. The consequent islands of automation call for the integration and interoperability of often scattered information architecture.

In order to make sense and use of large amounts of data, the meaning and the context of data is decisive in both the management of product definition, but also manufacturing operations management. The variety and complexity of data can be managed with sophisticated definition and utilisation of master data (Silvola, 2018). The master data have to be common between the operations so that the integrated approach of data management is enabled.

The concept of Closed-loop PLM was introduced by the PROMISE consortium. It is illustrated in Figure *23*, where dashed thick lines represent material flow along the product lifecycle including ‘recycling’ loops, while dotted lines represent information loops.

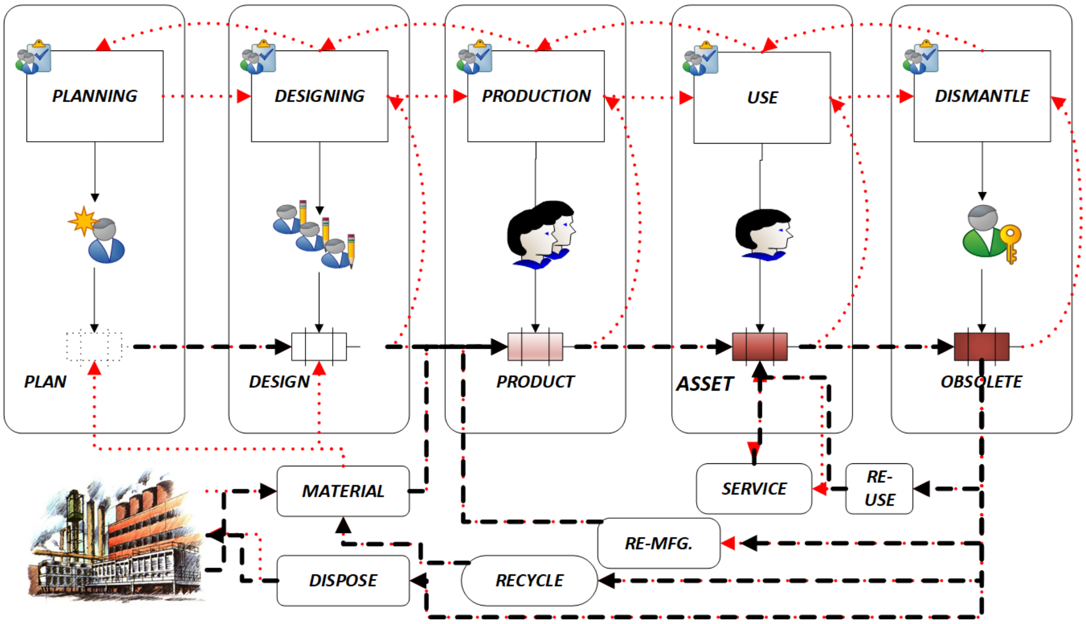


Figure 23. Closing the information loops (modified from Kiritsis et al., 2003, p. 194).

Originally, the feedback data was supposed to be collected with Product Embedded Information Devices (PEIDs), where barcodes or RFID tags indicate the identity of the product or item. With IoT the concept of Closed-loop PLM is partially realised. Many companies nowadays provide support services for augmenting their product offerings. However, full integration has not existed before the introduction of concepts such as digital thread and the utilisation of lifecycle data from a closed-loop PLM system in engineering design has begun only during the last few years when the concept of digital twin had received attention both in the offerings by vendors and customers' use cases.

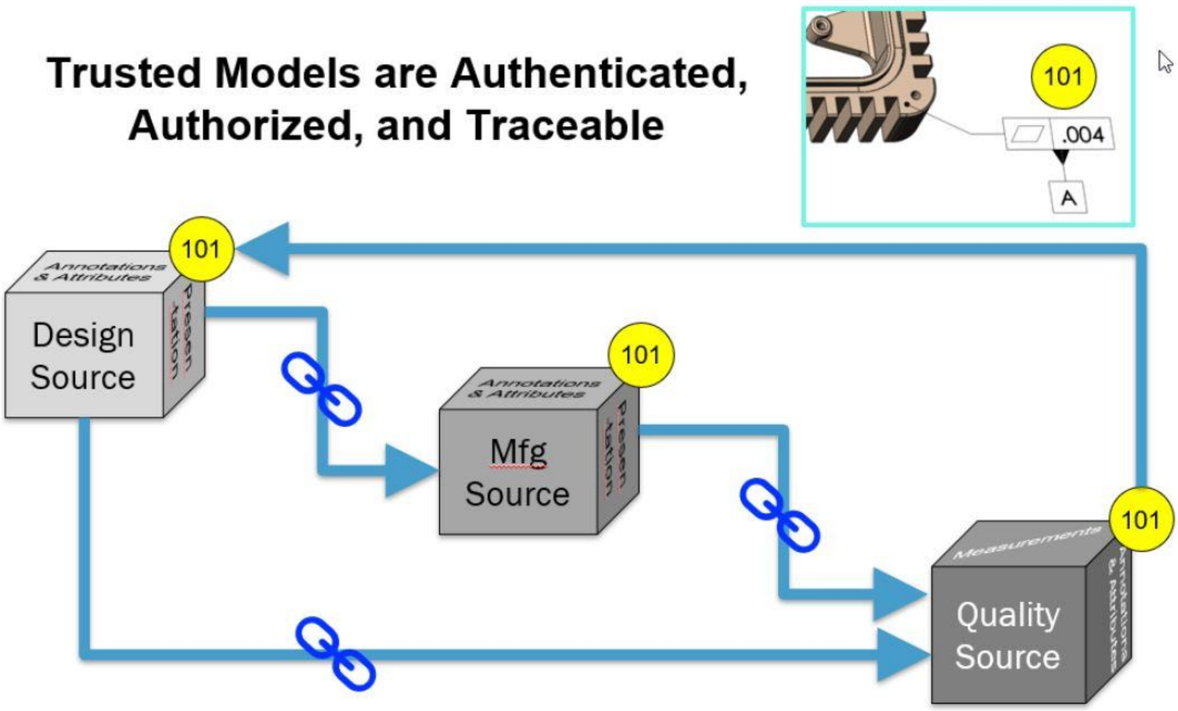


Figure 24. The closed loop of definition vs. verification by utilising QIF and unique identifier (adopted from Herron & Gelotte, 2019, p. 148).

One possibility to utilise MBD is to create a closed-loop from component definition to quality assurance and vice versa. Quality Information Framework (QIF) is an ANSI standard that "*defines an integrated set of information models which enable the effective exchange of metrology data throughout the entire manufacturing quality measurement process*" (ANSI/DMSC 2018, p. xxv). An illustration of closing the loop is seen in Figure *24*, where the unique identifier relates an attribute throughout the life cycle of a part and consequently enables the verification of a definition attribute. A similar benefit from the utilisation of MBD in engineering change management (ECM) process is presented by Quintana et al. (2012). The inadequate quality of change requests has been defined as a reason for prolonged ECM processes (Jokinen et al., 2017). A more precise definition of features and attributes with the MBD of items would obviously enable the refined description of a change request. One may think that elevating the closing of the loop to other levels and even to the product family stage would be comparable. However, the closed loop of the product variety definition would require different approaches and integrations.

6.2.2 Cyber-physical systems and manufacturing, architecture and 4.0

The ability to control and manage manufacturing tasks can evolve with the increased ability to harness and utilise data. Lee et al. (2015) have considered this as a stepwise development within the levels of smart connection, data-to-information conversion, cyber, cognition and configuration.

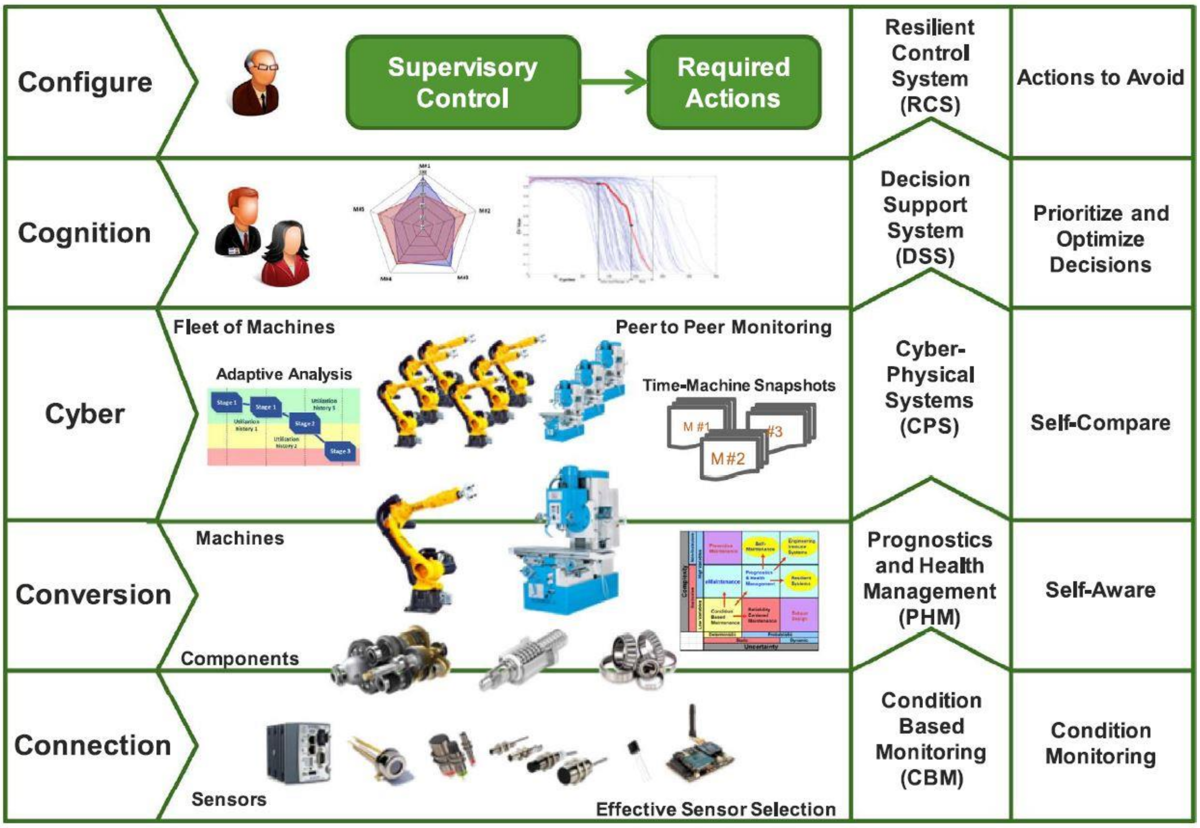


Figure 25. Applications and techniques associated with each level of the 5C architecture.

Lee et al. (2015) refer to the five-level CPS structure as a 5C architecture aimed as an aid for developing and harnessing the needed capabilities.

6.3 Supply chain management

To reduce uncertainty of increasingly dynamic supply and demand chains outside the main operations of companies, visibility of what is happening becomes key. Greater visibility increases performance and making decisions. However, to most companies the visibility of their relationships to their suppliers or other partners in the chain is limited or perhaps restricted to direct partners. Further upstream or downstream, the supply chain becomes increasingly unknown. In addition, data sources tend to be highly heterogeneous.

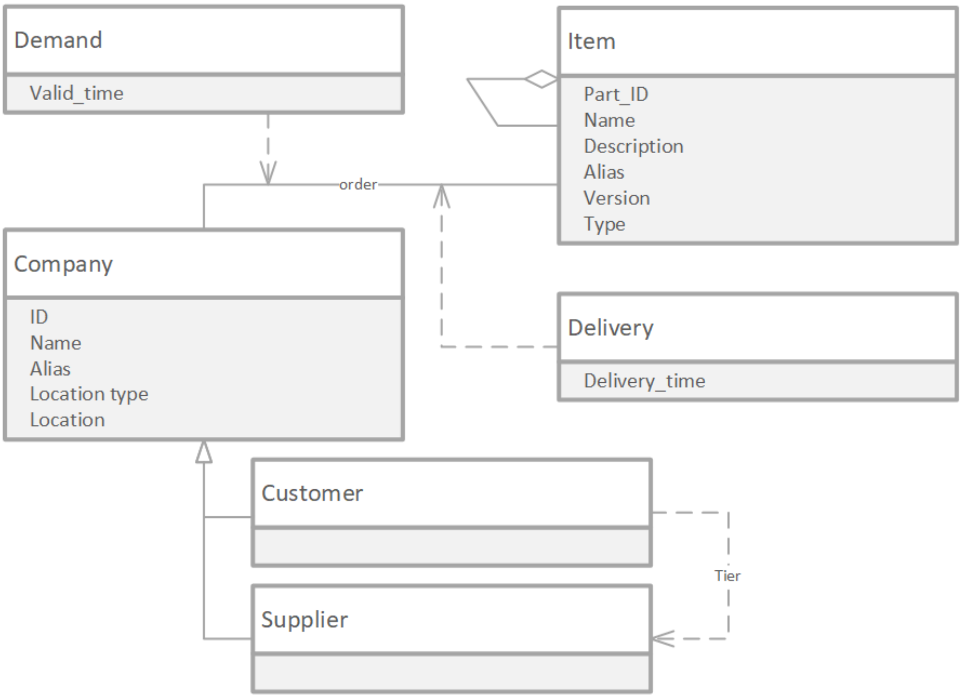


Figure 26 Conceptual representation of supply network structure data sources (modified from Zrenner et al., 2017, p. 122).

To increase understanding on missing connections between data sources for supply chain participants, the general knowledge on data sources and their meaning and suitability for all partners needs to be enhanced. Figure 26 shows an example of a conceptual representation of a supply network data structure based on the entity-relationship model. The entity-relationship model is used to depict the relationships of certain objects of interest, the entities, with possible attributes. It supports understanding the data foundation of business processes, in this case, supply chain network operations.

In this example, a given company in the supply chain has relationships with a specific supplier and customer, attributes such as company ID, name and location, and delivers products, which is likewise attributed with, for example, part ID, part name and part type.

The entities with their range of relationships and attributes are stored in a variety of data sources, such as company internal databases. However, not all data sources are suitable for all participants in a supply chain. In order to find out which data source is useful, a comparison can be made using a classification scheme as seen in the table below.

Table 1: Taxonomy of data sources (Zrenner et al., 2017, p. 124)

Dimension	Characteristics			
Data source availability	Internal	External – closed	External – open	
Data source interface	Internal interconnection	Traditional EDI	Web services	Offline data dump
Data source pricing model	Volume-driven	Time-driven	Unique	No
Data aggregation	Resource	Database	Record	Item
Data Occurrence/update	Stream	Event-driven batch	Time-driven batch	
Data ownership	One legal entity	Community	Public	
Data structure	Structured	Semi-structured	Unstructured	
Data format	Proprietary	Open		
Intra data standardisation	Value	Semantic	Syntax	No
Inter data standardisation	Value	Semantic	Syntax	No
Data currency	Forecast	Up-to-date	Outdated	
Data completeness	High	Medium	Low	
Data accuracy	High	Medium	Low	
Data sharing	proprietary	Free	Open	

Once a specific data source is taken for comparison, it is possible to go through the different dimensions to understand their relative characteristics and meaning for the respective supply chain participants. For example, the materials management database of a company may contain data on a product, its supplier in the market, its price, its current inventory and information on its consumption.

Thus, an analysis using the conceptual representation and the classification scheme supports identifying vital data sources, their description and comparison for the participants of a supply chain. This knowledge can subsequently be used to negotiate data source sharing in order to increase the visibility of the supply chain.

From the product quality point of view, taking into account the supplier's capability in relation to technical product requirements is essential already at the product development stage (Perttula, 2007). In Figure *27*, the focus is on the characteristics of orders and sales-delivery transactions. Instead, not only the order and the transaction data, but also data concerning the capabilities of a supplier play key roles in the developing supplier networks (see Figure *26*). The suppliers' capabilities in key performance areas (KPA), such as strategy, business models, processes, the use of performance indicators, efficiency in interfaces and information flow, are critical for strategic supplier development. Collecting and analysing the KPA data in a systematic manner enhances the strategic supply development in a company (Pulkkinen et al., 2019).

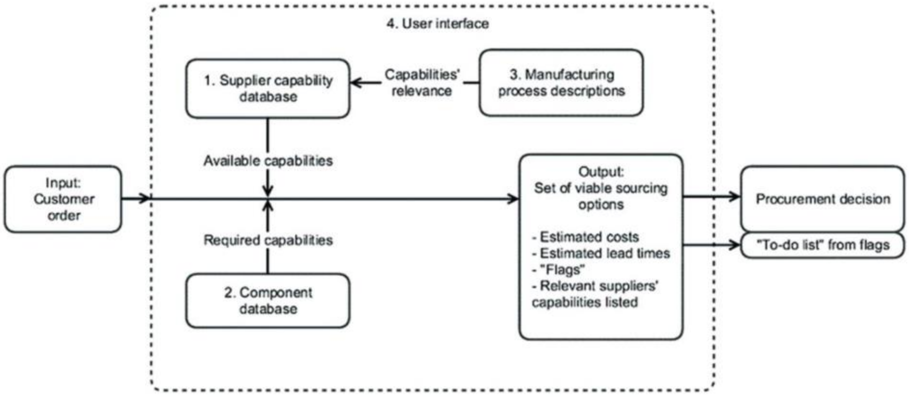


Figure 27. An overview of supplier selection tool by Hellström & Lagström (2017, p. 102).

In recent years, the ecosystem model for networks has emerged and the role of digitalisation within and among the partners of the supply networks (Pulkkinen et al., 2019) has risen. Digitalisation enriches supply networks with new opportunities for sharing data as well as integrating and co-ordinating operations. The visibility and timeliness of product and capability data have had positive effects on the quality (Perttula, 2007), responsiveness and productivity of suppliers (Pulkkinen et al., 2019).

6.4 Production activities functional model and data flows

Production activities include the actual manufacturing processes that make the products and ensure their quality, together with all the management and preparation activities that support the efficient use of these processes. Production activities determine the products to be manufactured at any one time, the order in which they are produced, and the allocation of resources to their production. Production activities include:

- Physical shop-floor processes, including machining, sheet metal stamping, die casting and injection moulding, finishing processes, inspection processes and mechanical assembly
- Materials requirements planning and manufacturing resource planning
- Production scheduling and control
- Tool management
- Inventory control
- Equipment maintenance and human resource management

These activities create and are controlled by information, data flows.
Figure 28 shows functional elements of the reference architecture for smart manufacturing. All the functional elements need data management and transfer between elements, from product design to the engineering phase for manufacture and planning of production systems. The Produce Products-phase (
Figure 28) comprises the actual work, all the customer order-related data on material, processes and scheduling of activities and monitoring data from actual process.

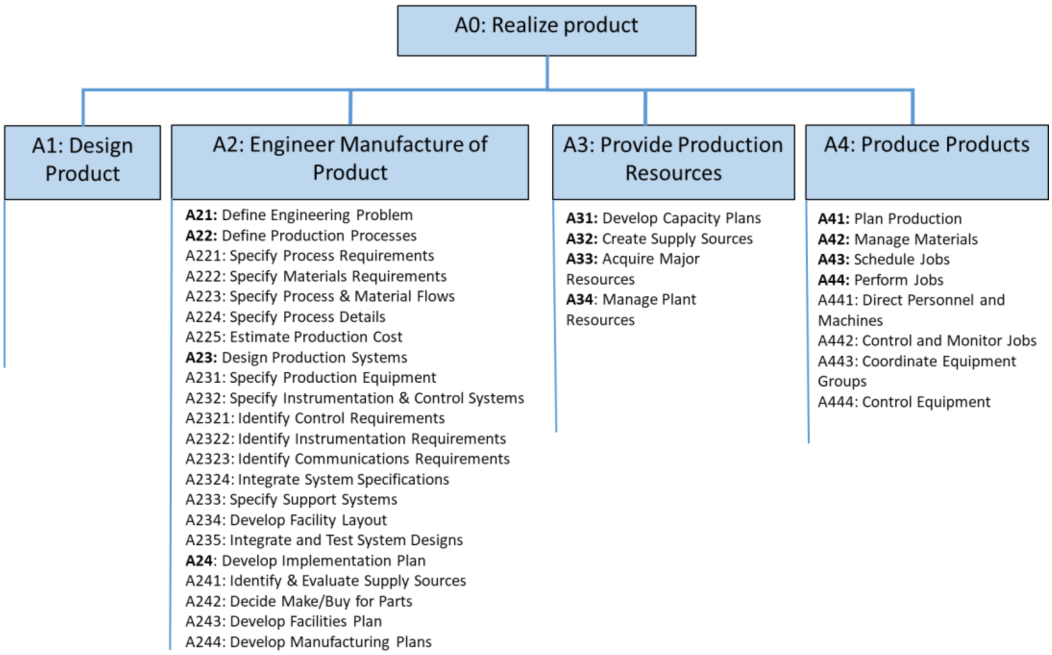


Figure 28 NIST Reference architecture for smart manufacturing functional model elements for realised product (Barkmeyer and Wallace, 2016).

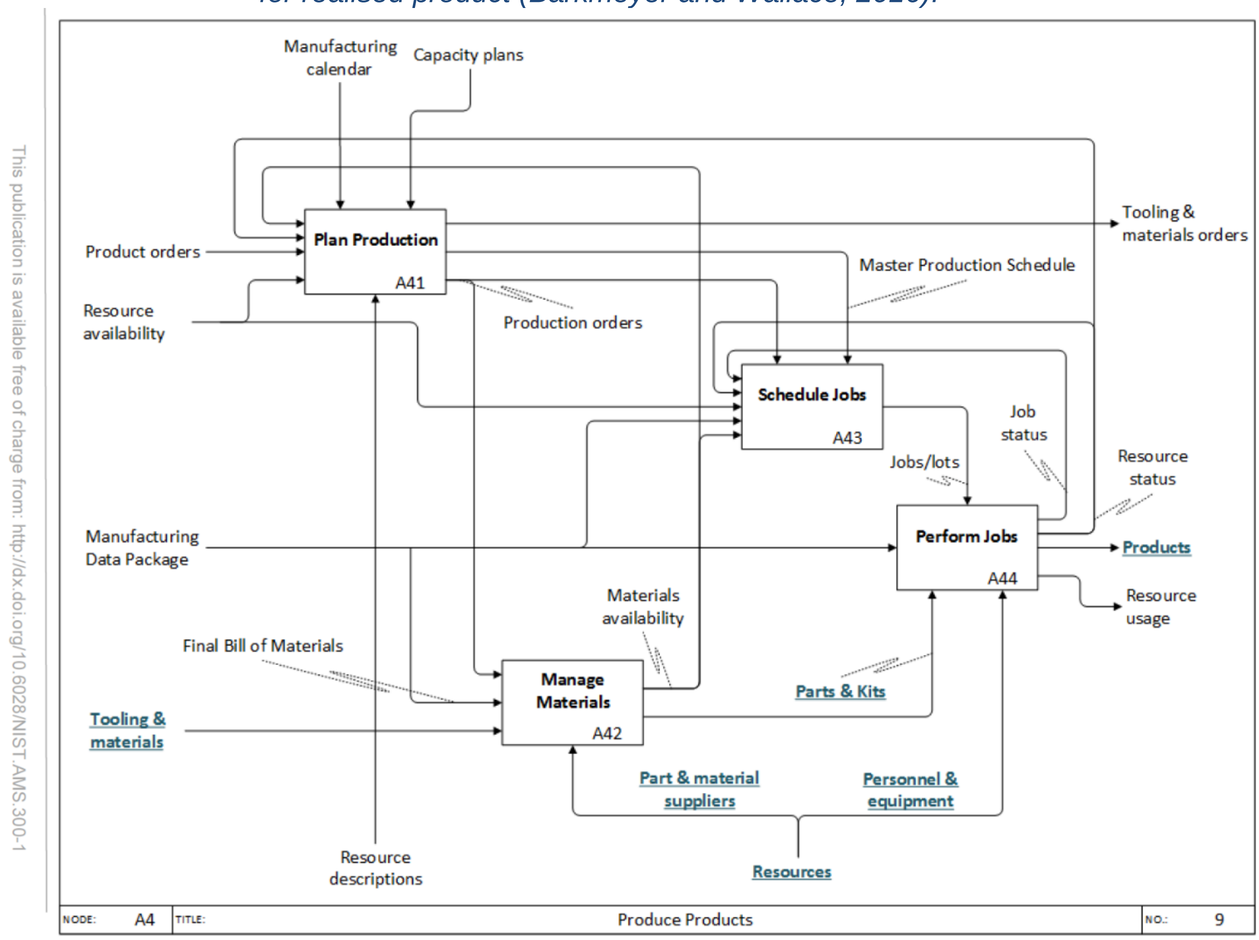


Figure 29 Produce products IDEF0 functional model shows data and information flow (Barkmeyer and Wallace, 2016).

Figure 29 shows an IDEF0 model of the function, produce products; more details are found in (Barkmeyer and Wallace, 2016).

ISA-95 is the international standard for the integration of enterprise and control systems. ISA-95 consists of models and terminology. Its official name is "ANSI/ISA-95 Enterprise-Control System Integration", known internationally as IEC/ISO 62264. ISA-95 defines in detail an abstract model of the enterprise, including manufacturing control functions and business functions, and its information exchange. It establishes common terminology for the description and understanding of enterprise, including manufacturing control functions and business process functions, and its information exchange. It defines electronic information exchange between the manufacturing control functions and other enterprise functions including data models and exchange definitions (https://www.isa.org/isa95/).

Functions for enterprise production control and flows based on ISA-95 (IEC 622264 Enterprise-control system integration) in Figure 30 include:

- Order processing (1.0)
- Production scheduling (2.0)
  - Production control
  - Process support engineering
  - Operations control
- Operations planning (3.0)
- Material and energy control (4.0)
- Procurement (5.0)
- Quality assurance (6.0)
- Product inventory control (7.0)
- Product cost accounting (8.0)
- Product shipping administration (9.0)
- Maintenance management (10.0)

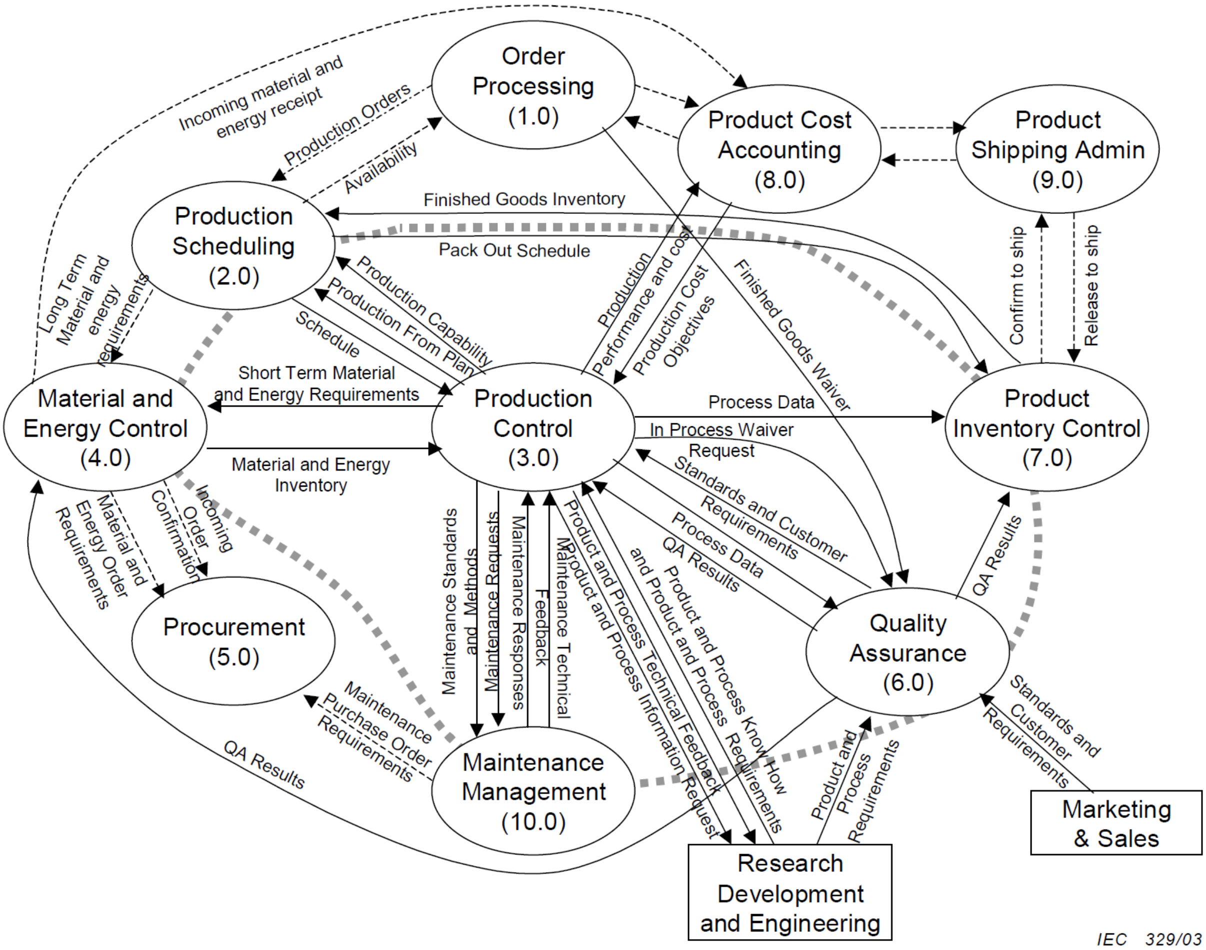


Figure 30. Functional enterprise/control model shows the functions and information flows (IEC 62264).

Figure 31 depicts the different levels of a functional hierarchy model: business planning and logistics, manufacturing operations and control, and batch, continuous, or discrete control. The model shows the hierarchical levels at which decisions are made. The interface addressed in the standard IEC 62264-1 Enterprise-control system integration – Part 1: Models and terminology is between Level 4 and Level 3 of the hierarchy model. This is generally the interface between plant production scheduling and operation management and plant floor coordination.

Table 2 shows the ISA-95 levels and examples of related Information Technology and Operational Technology systems.

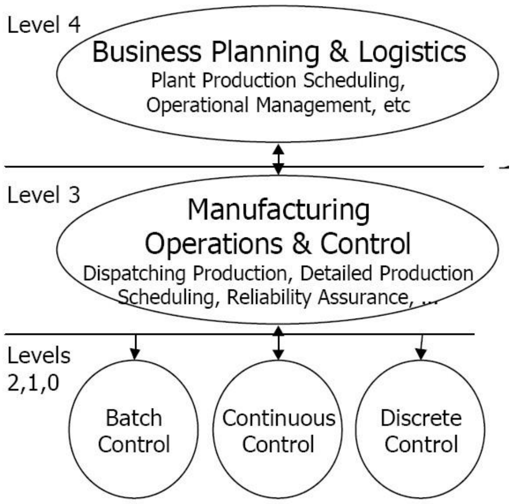


Figure 31 Simplified version of the Purdue Hierarchy Model – Functional Hierarchy.

Table 2 ISA-95 levels

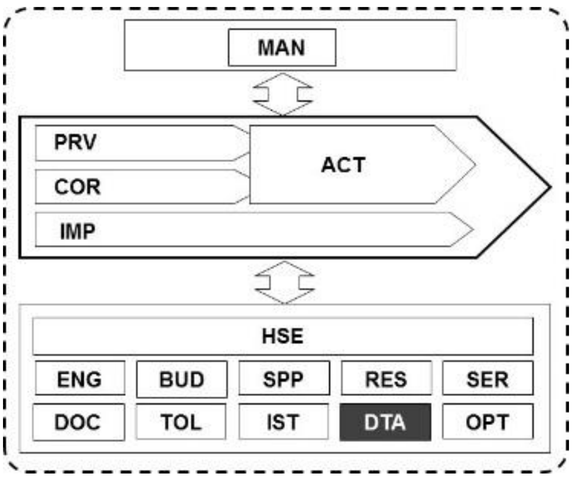
Purdue Hierarchy Model – ISA-95 levels	Information Technology, Operative Technology systems
Level 4: Business and logistics systems — Managing the business-related activities of the manufacturing operation. ERP is the primary system; establishes the basic plant production schedule, material use, shipping and inventory levels.	ERP, SCM Time frame: Months, weeks, days, shifts
Level 3: Manufacturing operations systems — Managing production work flow to produce the desired products. Batch management; manufacturing execution/operations management systems (MES/MOMS); laboratory, maintenance and plant performance management systems; data historians and related middleware.	MES, MOM Time frame: shifts, hours, minutes, seconds.
Level 2: Control systems — Supervising, monitoring and controlling the physical processes. Real-time controls and software; DCS, human-machine interface (HMI); supervisory control and data acquisition (SCADA) software.	DCS, HMI, SCADA
Level 1: Intelligent devices — Sensing and manipulating the physical processes. Process sensors, analysers, actuators and related instrumentation.	PLC
Level 0: The physical actual processes	Sensors & Signals

6.5 Asset management and life cycle services

Asset management covers the activities in operation, maintenance and improvement of assets that are required for optimal life cycle management and economic sustainability. ISO standard ISO 55 001 also states that the organisation should determine the information needs related to its assets, asset management and its asset management systems. Issues to be considered include e.g.:

- The value of the information to enable decision making and its quality relative to the cost and complexity of collecting, processing, managing and sustaining the information;
- The participation of relevant stakeholders to determine the types of information required to support decision making as well as to ensure the completeness, accuracy and integrity of the necessary information;
- The alignment of the information requirements for different levels and functions within the organisation;
- The establishment of data collection processes from internal and external stakeholders (including contracted service providers); and
- The data flow and integration of information sources for planning, operational and reporting technology systems, appropriate for the size, complexity and capability of the organisation.

The role of the data is recognised and defined also in other asset management-related standards as a crucial element of successful operation. As an example, the recent maintenance process standard, EN 17007: 2017, involves a sub-process for managing data (Figure 32).



- Store and validate the raw data in a library and/or a database
- Evaluate the reliability and maintainability of the items by maintaining an actual state assessment of the items
- Draw up and maintain an up-to-date list of critical items
- Evaluate and analyse maintenance data and HSE data
- Evaluate and analyse data related to spare parts
- Evaluate and analyse cases of known or predictable obsolescence
- Collect and analyse events at other organisations
- Compare maintenance practices and materials used by other operators or recommended by suppliers
- Monitor methods, technologies, regulations, standards, etc.
- Save and provide access to data in a data processing system
- Calculate, save and provide access to performance and monitoring indicators

Figure 32: Maintenance process and data management tasks (EN 17007)

Companies providing hardware and services for the operations and maintenance phase (O&M) could offer value to the customers by being able to provide knowledge and consequent actions or action plans as a service instead of mere data collecting and sharing. An example of such development is given in Figure 33. In this case, the service provider is capable of refining the data to the information and knowledge that supports the decision-making of customers (asset owners) to improve their business. Decision situations that could be supported include daily operations, developing the assets and operations, or with long-term strategic investments, and a wide variety of other technical issues and business situations (Kortelainen et al., 2017).

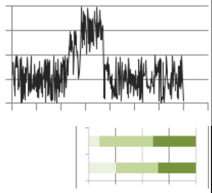
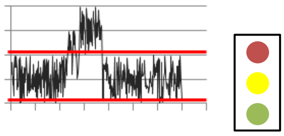
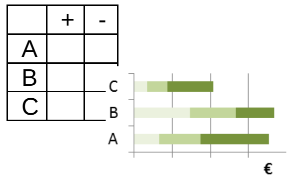
Data as a Service	Information as a Service	Knowledge as a Service	Wisdom as a service
			
Database containing data on measurements, maintenance history, etc.	**Report** representing various KPIs and visualisations based on the gathered data	**Interpretation of information** by answering, for example, the questions: • When are the measurement values deviating? • When does the maintenance programme need to be changed? • When does a need for technological changes in the fleet exist • etc.	**Evaluating of alternatives** by answering, for example, the questions: • How are deviations dealt with? • How should the maintenance programme be developed. • etc.

Figure 33: DIKW hierarchy applied to the maintenance service levels (Translated from Kunttu et al., 2017).

Industrial assets are becoming increasingly IoT-instrumented, intelligent and connected. Due to the increasing expectations from technological advances and digitalisation, future production systems will be probably even more complex than existing ones (Alcacer & Cruz-Machado, 2019). Also, working environments will give much more importance to the human dimension, making the most of worker's knowledge, cultural background, autonomy and independent decision-making. However, at the same time responding to worker's needs to support work tasks and processes, leaning activities towards the maintenance of required skills (Mohamed, 2018). The workplaces of the future will be based on methodologies for enhancing flexible, safe and data-driven smart production, where adequate levels of automation are applied, while maintaining a level of employment with highly satisfied and skilled workers.

In the future, ICT companies will be key players in future ecosystems as platform operators and will be essential for all knowledge-based services of machinery and service providers (Kortelainen et al., 2019a). Such collaboration schemes have been announced by global companies like KONE and IBM, and ABB and Microsoft. Machinery suppliers offering asset life cycle services may have access to their global product fleet (assets) and they can collect data, monitor and even operate or control their fleet in different geographic locations over their own or shared platforms. According to Davies (2004), the competitive advantage is not simply about providing services, but how services are combined with products to provide high value "integrated solutions" that address a customer´s business or operational needs. The service providers have to develop the excellence in refining data in a way that delivers more value to the asset owner.

These trends will require new data-driven assisting service and proactivity-supporting technologies for an optimised use of workers' knowledge, cognitive and creative capabilities. The machine-readable data converted to proactive knowledge will constitute the intelligence of these technologies. The core proactive knowledge will be co-constructed within the product during the design phase, while other parts are gathered during the product life cycle using sensing and communication capabilities embedded in the smart product systems.

The data which is turned into proactive knowledge can be captured from a variety of different media such as design drawings, 3D models, manuals, operational and maintenance data, and company internal and external platforms and internet. Accordingly, supporting technology such as real-time contextual data processing and reasoning, as well as an efficient capture of domain- and processes-related knowledge at design time as well as making them adaptable and extendable at run-time, are required to support humans in the decision-making process.

There are a number of technologies such as reasoning engines and ML algorithms (Wang et al., 2018; Hou et al., 2017) as well as domain-specific ontologies and technology developed based on Linked Data principles (Niskanen, Kantororovitch, & Golenzer, 2012; Schabus &Scholz, 2017; Giustozzia et al., 2018) that can be exploited to enable selection of appropriate interaction mechanisms and a subset of knowledge, that is relevant for a given context, and which should be exchanged with another product or system and/or with a user to support decision-making process.

The growing role of data and AI in business decision making is well recognised. With a significant amount of data, an ultimate AI system creates a digital model of the organisation. Based on information concerning input (such as people and assets) and achieved results, patterns can be identified and recommendations and actionable insights for smarter decisions can be made. For that to happen, numerous challenges related to the quality of data and its accuracy, as well as the ability to reuse ML algorithms for different scenarios and/or changing business goals and less generic situations (that are not included in the historical datasets), need to be addressed (BDVA, 2018).

7. Data interoperability and standardisation

7.1 Reference Architectures for Interoperable Manufacturing

A number of reference models and architectures have been developed in recent years to address the issues of integration and interoperability in the context of smart manufacturing. According to Barkmeyer and Wallace (2016), the meaning of reference architecture is:

- identification of the functions required to accomplish a set of objectives in a given domain;
- identification of types of systems (components) that perform, or support human agents in performing the activities that implement those functions; and
- identification of the nature and content of the interfaces required among those systems.

The following is a list of selected reference architectures for smart manufacturing.

PERA, Purdue Enterprise Reference Architecture, is a 1990s reference model for enterprise architecture, also known as ISA95. It provides a model for enterprise control, which end users, integrators and vendors can share in integrating applications at key layers in the enterprise. Enterprise control is the ability to combine control, intelligence and process management to enable business optimisation that is inclusive of business and production operations. It combines the strength of both business processes and production operations processes. It is the deliberate act of synchronizing business strategy with operational execution in real-time to enable closed loop business control across an enterprise. The PERA 95 model is used in several ISO standardisations, e.g. a series of IEC 62264 Enterprise-control system integration.

SIMA, Systems Integration of Manufacturing Applications, is a reference architecture developed at The National Institute of Standards and Technology (NIST), (Barkmeyer, 1996) that addresses product design engineering, manufacturing engineering, production systems engineering, and production activities, corresponding to the four top-level activities: (1) Design Product, (2) Engineer Manufacture of Product, (3) Engineer Production System, and (4) Develop Products. The management of Engineering Workflow is also included (See Figure *30*).

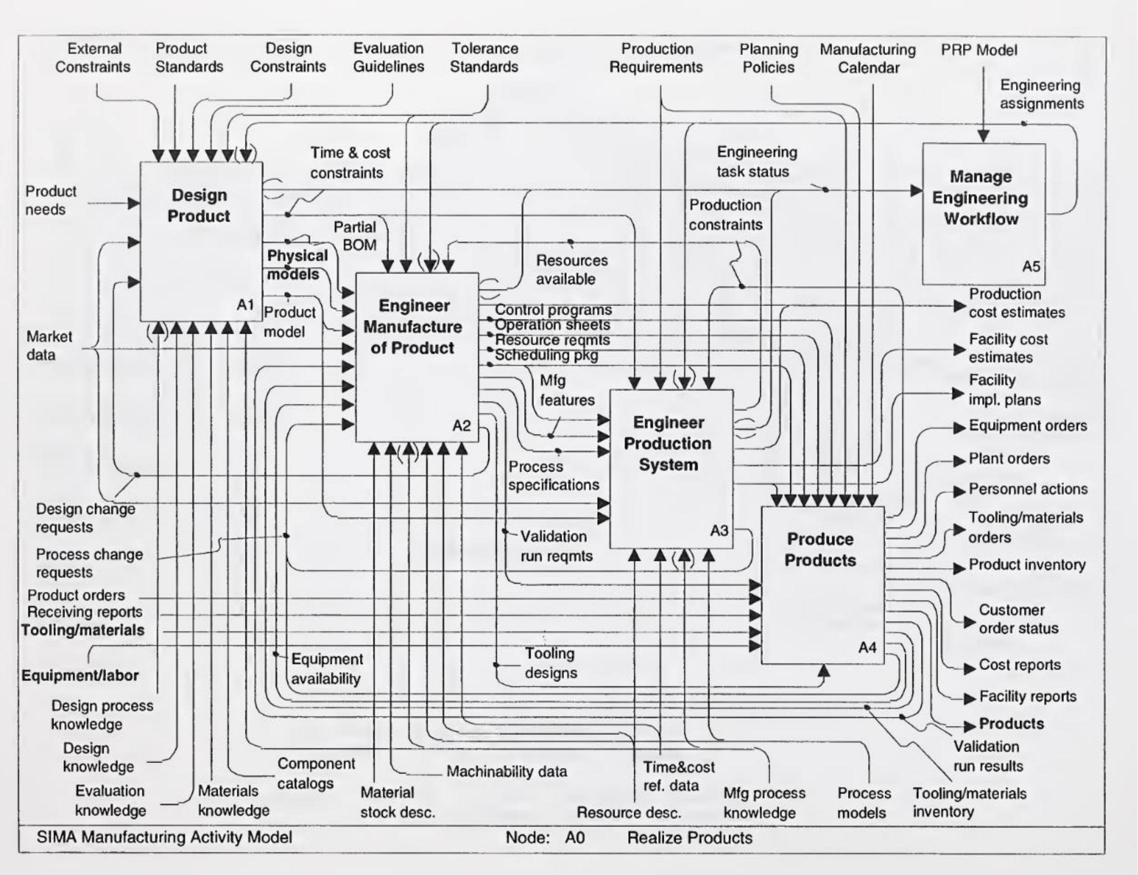


Figure 34. *NIST SIMA Reference Architecture Realise product IDEF0 activity model.*

NIST Reference Architecture for Smart Manufacturing, Functional Models are shown in Barkmeyer and Wallace (2016). The main activities of the SIMA architecture are consistent with recent developments in the IoT reference models, such as the German Reference Architecture Model Industry 4.0 (RAMI 4.0) and work from the Industrial Internet Consortium (IIC), IIRA.

RAMI 4.0, Industry 4.0 reference architecture model, can be seen in Figure 35.

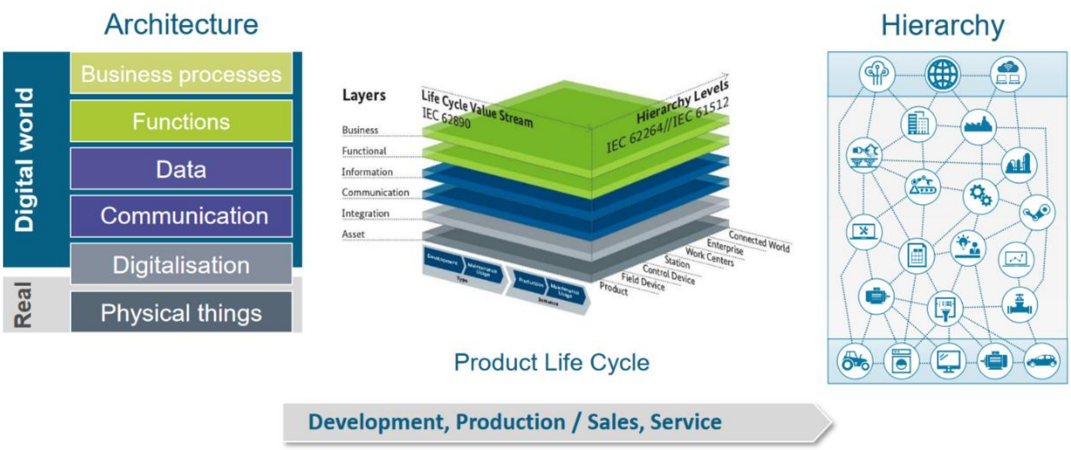


Figure 35. RAMI 4.0 covers the product life-cycle, the business aspects and the factory hierarchy.

RAMI 4.0 is a three-dimensional layer model that compares the life cycles of products, factories, machinery or orders with the hierarchy levels of Industry 4.0. More information on RAMI is available at https://www.plattform-i40.de/PI40/Redaktion/EN/Downloads/Publikation/rami40-an-introduction.html.

IDS, International Data Space, is a reference architecture model that can be seen at: https://www.internationaldataspaces.org/.

The specification of the IDS Association forms the basis for a data marketplace based on European values, i.e. data privacy and security, equal opportunities through a federated design, and ensuring data sovereignty for the creator of the data and trust among participants. It forms the strategic link between the creation of data in the IoT on the one hand and the use of this data in machine learning (ML) and artificial intelligence (AI) algorithms on the other hand.

Digital responsibility is evolving from a hygiene factor to a key differentiator and source of competitive advantage. Future data platforms and markets will be built on design principles that go beyond our traditional understanding of cybersecurity and privacy. Based on strong data ethics principles, the IDS Reference Architecture Model puts the user at its centre to ensure trustworthiness in ecosystems and sovereignty over data in the digital age as its key value proposition.

IDSA defines a reference architecture, which supports sovereign exchange and sharing of data between partners independent from their size and financial power. Thus, it meets the needs of both large and small and medium enterprises (SMEs). Further down the road, it may be taken up by individuals as well. Whether or not IoT device data is concerned, in on-premise systems or cloud platforms, the IDSA aims at providing the standard for sharing data between different endpoints while ensuring data sovereignty. IDS Reference Architecture Model 3.0 is available at https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf.

IIRA, Industrial Internet Reference Architecture (IIRA) model, was developed by the Industrial Internet consortium (IIC) based on the ISO/IEC/IEEE 42010:2011 standard. More information about IIRA is available at https://www.iiconsortium.org/IIRA.htm.

The organisations behind RAMI, IDSA and IIRA do collaborate.

7.2 Standardisation for smart manufacturing

The purpose of standardisation is typically to improve the interoperability and compatibility of, e.g., products, systems, technologies, methods or data. Standardisation serves the end users by, e.g., guaranteeing compatibility and interoperability of solutions from different vendors. It also enables solution providers the increased technical stability for developing new technologies and building solutions. Especially for small companies and organisations, standardisation can provide interfaces that enable investments that would otherwise be too risky.

The smart manufacturing standards framework includes the following several main parts (Qing et.al., 2018):

- Smart design standards: the group of standards are expanded along the order of design activities, supported by data management standards. The standard framework decomposition does not follow the classification of design subjects.

- Smart production standards: the group of standards are expanded based on working process and technical support.

- Business operation and management standards: the group of standards are focused on management activities for design and production. Commercial applications like ERP, SCR, CRM, MES, are not used as standards categories. Their implementation standards are discussed in the combined management standards group.

- System integration standards: this group of standards relates to common technologies that integrate systems of different domains.

- Fundamental technologies and supporting environment standards: the group of standards includes standards on common supporting technologies, such as infrastructure, database, meta data technology and so forth.

Standardisation is an important factor for the whole manufacturing industry. There are numerous standards for different purposes in the domain and there are activities to share the awareness of the importance of standards and envision how to strengthen the domain also via standardisation. More information on smart manufacturing standardisation is available at Industry 4.0 standardisation roadmap version 3 (DIN e.V., 2018), NIST smart manufacturing standard landscape report (Yan Lu et al., 2016) and ISO Technical Report, the big picture of standards (ISO/TR 23087:2018).

Sometimes the application field is too complex and evolves so rapidly that standardisation is not able to keep up. In the design and engineering of products or systems, the challenge is in the complexity of the process, complexity of the data and its representation and, increasingly, in the volume of data. The wide variety of computational methods and large number of simulation and analysis tools implementing them, and the wide offering of engineering and design systems, is one of the root causes of the challenge. Some standardisation of data formats exists, but standards do not cover all the needs and cannot be kept synchronised with the progress of methods and tools. An example of data exchange and integration challenges is computer-aided design (CAD) in mechanical design. CAD tools are used for defining design plan models and to produce design documents. The design models can be done in 2 dimensions (2D) or in 3 dimensions (3D). The 2D design tools are in practice sophisticated 2D drawing software applications. In 2D design models, graphical elements, such as dots, lines, arches, circles and curves, are used for representing either the design, i.e. the shape of the designed object, or the meta-data of the design, such as dimensions of or notes concerning the design. This means that, in principle, identical graphical elements may have different meaning, i.e. semantics.

When transforming the CAD model data from one software application to another, the data may be very difficult to transform so that the semantics of the elements is preserved. For example, a dimensioning line in the source document will be treated as a dimensioning line in the target document, not as a geometry line. In 3D, the challenge is even more complicated. The 3D geometry information can be represented in several ways. The main approaches are constructive solid geometry (CSG), which represents the complex geometries by a set of Boolean operations (set theory) of a set of primitive geometries, such as boxes, spheres and cylinders, and boundary representation, which represents the complex geometries with closed mathematics surface patches that form a closed volume (in the case of solid geometry). In the case of boundary representation, the surface patches may have different mathematical representations, such as 3D Bezier surfaces or non-uniform rational B-splines (NURBS). When the data is translated from one CAD tool to another and the internal model data representations differ due to the differences in the way geometry information is managed and processed in the software application, the data transformation may be challenging. For 2D and 3D geometry CAD data, there are several standards and de facto standards, such as ISO 10303-242, Initial Graphics Exchange Specification, IGES (US PRO, 2006), and Autodesk AutoCAD DXF (Autodesk, 2013).

The ongoing trend in design and engineering is that the large design and engineering solution providers offer integrated solutions for most of the tasks in the design and engineering process. While the issue in data interoperability in this approach is usually solved, the end-user companies are tied into a solution from one vendor. This so-called vendor locking situation may be a risk for the end-user company, as it increases the dependency of the solution provider, as the data content, in the form of design and engineering outcome, is tied to the specific solution. One aspect that can prevent vendor locking has for a long time been standardisation. When the information of different design and engineering tasks is managed in standardised form, the information content is not tied to specific tools and can be processed and utilised with any standard compliant tool. Standardisation can be seen as a way to divide the information domain into smaller pieces, enabling the information to be processed with any compatible tool. This approach leaves the end-user the ability to select the tool that best fits his/her needs. In addition, it enables competition between the solution providers, which usually improves the value for the end-user. It also enables new and smaller solution providers to enter the market with their solutions.

8. Summary and future research topics

8.1 Summary

Hyper-Agile Cognitive Industry refers to the intelligent networking of machines, processes and humans for industry with the help of information and communication technology. To enable agile manufacturing, product development, production plants, supply value networks and logistic systems, design has to be flexible and reconfigurable on the fly to respond quickly to customer needs, production uncertainty, and market changes.

The future of manufacturing lies in being cognitive, smart – capable of agilely adapting to a wide variety of changing conditions. It is essential to see the full value chain, which includes suppliers and the origins of the materials and components needed for various forms of smart manufacturing, the end-to-end digital supply chain and the final destination of all manufacturing, regardless of the number of intermediary steps and players: the end customer. Automation, robotics and human technology interaction are the key aspects of achieving customer responsiveness, productivity and sustainability.

In the value network, cognitive technologies look deeply into design data, the manufacturing process, business environment and product usage to derive information that has tangible value for a manufacturer. Cognitive manufacturing is all about exploiting data from diverse, heterogeneous sources, structured as well as unstructured data and applying advanced analytical models and process to create a knowledgeable system that is continuously learning. It is able to make insightful operational recommendations for manufacturing based on a comprehensive understanding of that data, and events behind of data.

Data as a relatively simple concept has been shown to be challenging and extremely complex. The rapid change of the role of data in industry and business, and especially the new and rising opportunities to have new business and services based on data, have emphasised also the challenges in utilising data. From one point of view, data is seen as the new valuable raw material for business. However, from another point of view, it is an important enabler for remarkable improvements in industry and the essential core of the Industry 4.0 vision. As for any central resource in business, there are technical, but also many other aspect in using data, such as security, openness and availability, processes and even political issues.

8.2 Future research topics

8.2.1 Vendor locks

One of the biggest concerns for companies who have moved to the cloud is the vendor lock-in. Vendor lock-in occurs when a customer becomes dependent on a vendor for products and services (Druzin, 2016). Whether the lock-in is directly forced by the vendor or is caused by complex technical dependencies (such as different data formats) that cannot be undone and tying future success to one vendor represents a risk for any organisation (Opara-Martins et al., 2016). To avoid the vendor lock-in situation, many organisations are adopting a so-called multi-cloud strategy, in which companies deploy applications across different clouds, using a variety of cloud partners. However, the multi-cloud strategy also brings challenges in terms of data integration and interoperability, for example (Chauhan et al., 2018).

8.2.2 Data as a digital asset

With the growing significance of data as a key component of doing business, companies are treating data more and more as an asset that delivers or has a potential to deliver value that can be monetised (Fleckenstein & Fellows, 2018, p. 11). The recent BDVA Position paper (de Vallejo et al., 2019) raises the business models that can exploit the value of data assets as one of the great opportunities in the next 10 years. The BDVA Position paper also emphasises that the most innovative data-driven business models are showing a wide variety of value creation possibilities, from direct data monetisation to access-based valorisation of data assets on sharing platforms. Subjectivity and case-specificity applies also to the value of data that is strongly linked to its use case and context (e.g. Koutroumpis and Leiponen, 2013; Laitila, 2017).

Guidelines for data pricing models and monetisation are needed and new forms of value creation uncovered by new sharing mechanisms need to be explored (Gatteneo et al., 2019; de Vallejo et al., 2019). In the future, data and data-intensive services like digital twins could be regarded as digital assets that should be created, developed, managed and valuated with the same scrutiny as physical assets.

8.2.3 Refining and sharing data in an ecosystem

Manufacturing companies operate in complex ecosystems with multiple actors who could benefit from the data. The digital strategy requires also sufficient investment and commitment from an organisation (Wixom and Ross, 2017; Laney et al., 2015). Many of the costs related to the design, implementation, deployment and operation of IoT services are unknown or insufficiently assessed (Nicolescu et al., 2018*)*. In addition to the ICT infrastructure, the costs of developing the skills and capabilities needed to cover all stages of the data value creation process (Figure *14*) have to be considered and assessed with the incurring benefits and data monetisation options. Also, the companies are specializing in different roles in the data value chains (Spijker, 2014). In industrial ecosystems, the main barrier to sharing data seems to be the strong sense of data ownership. Issues like unclear values (transparency vs. privacy), lack of knowledge about handling data, security and lack of understanding the potential of the data create further barriers (Kortelainen et al., 2019b). Many questions dealing with data ownership, security and principles of sharing costs and profits in a business network of an ecosystem lack commonly accepted solutions.

8.2.4 The lifetime of the data

Like physical assets, also data assets have their lifetime and obsolete data have diminishing or no value to an organisation. In data management, the cost of storage is relatively low, though storing large amounts of data online can be expensive. Removal of obsolete data results in more efficient searches, data integration and reporting (Fleckenstein & Fellows, 2018, p. 20). In a manufacturing site, the data collected from the operation and maintenance of a machine becomes obsolete, e.g., if the item undergoes a major refurbishment, the operational mode is changed in a crucial way, or the machine is replaced with a different one. The lifetime and relevance of the data becomes also a major concern in machine learning and AI solutions. If the physical system, operational environment or other conditions change, the algorithms or teaching material may no longer be valid. Thus, development of data life cycle management and methods to assure the relevance and accuracy of the data are more important than ever.

Challenges arise from a low proportion of abnormal situations. For example, if data is used for condition monitoring and data-based technologies, such as artificial neural networks (ANN) and machine learning (ML), are used for fault detection, the challenge may be the lack of data of fault situations for ANN training. One work around for this is to use simulated data in the ANN training. An accurate enough simulation model can produce data of different kinds of fault situations and the model can be changed so that there is enough variety in the results for reliable ANN training.

In industry, sensors and automation systems collect a huge amount of data. However, a great portion of this data arises from normal operation and repeated data sets have low information value. Abnormal situations like production disturbances and failures happen seldom in a statistical sense. In addition, the causes and impacts of these unexpected and undesired events are various. This fact raises a challenge for developing machine learning and AI algorithms, as the time and effort needed to create teaching material increases rapidly.

8.2.5 Safety and security of AI in manufacturing

The introduction of AI technologies brings several benefits to manufacturing, for example in decision-making support, predictive maintenance and optimisation of operations. To be able to deploy AI technologies and achieve a sufficient level of user acceptance, the safety and security of AI applications needs to be ensured. At the moment, testing and validation of AI systems are major challenges, and several concrete safety problems hinder the deployment of advanced AI systems. A systemic perspective is needed to address the safety of AI applications and especially AI-human co-operation in the manufacturing industry.

8.2.6 Automated data interoperability

Data is exchanged increasingly between information systems and software applications. Most of the new data integrations require manual coding and changes to the software applications and systems. Even though there are data modelling and data transformation technologies, such as the EXPRESS language, the Semantic Web technologies, and XML and related technologies, data integration is not generally automatic. IoT as a framework and a concept, and its technologies, provide one example platform for improved data interoperability, but they do not solve the general challenge. If data exchange technologies were at the same level as, e.g., USB is in computers, it could revolutionise the use of data, both in industrial as well as in consumer markets.

8.2.7 Knowledge management as a business asset

There has been discussion about tacit knowledge in the industrial context for decades, but now a general solution has been provided. The management of expert knowledge has been found to be especially challenging due to its complexity and variety. Many attempts to utilize AI have failed with consistency problems. The contents of knowledge bases require version and variant management functionalities in a similar way to data management systems. Formal knowledge management, combined with general data management, may become a valuable tool for companies to keep their valuable know-how and knowledge within the organisation and up-to-date. If the knowledge is maintained in a formal form, it can serve as the source of intelligent systems and enable improved operating process efficiency as well as new services and solutions.

8.2.8 Accuracy, reliability and traceability of the data

With increasing value, data also becomes vulnerable to fraud. As an example, imagine a scenario where a satellite launch fails due to a falsified material certificate, and the material used for the launch turns out to be fake aluminium. Some material suppliers have already established a platform where individual product batches can be traced. However, in the dawn of the data economy, the significance of the accuracy, reliability and traceability of the data have to be highlighted and methods to assure data quality have to be developed.

8.2.9 Relating customisation and agility to economy for decision-making

The large amounts of data do have an effect on life-cycle operations of products, such as manufacturing. For example, the rise of item numbers can lead to an increase of tools and equipment, unnecessary variation in production and increase of set ups and changes in production, large stock of different kinds of parts, facilities and more-specific logistics. These consequences increase overhead costs and have often a poor effect on profitability. For companies, integrating product data with economy-related data, especially from enterprise management, manufacturing operations and supply chain management systems, creates a potential for understanding the profitability of products. The ability to eradicate unprofitable products and variants is a necessity for industry. Research should provide models and methods for integrating data from different sources for data-driven management of products and operations.

8.2.10 Data in circular manufacturing

One of the key obstacles to moving the circular economy is the data that does not exit or stays in silos. On the other hand, data markets would open up new opportunities to create businesses related to, e.g., reuse, repair, remanufacturing and recycling of products. Information on products' condition, materials, use history and lifecycle stage (traceability) enables increased product circulation, demand prediction, predictive maintenance and intelligent asset inventory management. Remanufacturing reguires information on the versions of assemblies and especially on the interface characteristics of the modules to be re-manufactured. The compatibility of old and new versions requires the compatibility of data. Research is needed to recognise data needs and sources, and to identify data semantics and formats to serve remanufacturing, refurbishment and, in general, transition towards the circular economy.

References

[1] Abiteboul, S., Cluet, S., Milo, T., Mogilevsky, P., Siméon, J., & Zohar, S. (1999). Tools for data translation and integration. *IEEE Data Eng. Bull.*, Vol. 22(1), pp. 3–8.

[2] Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, Vol. 16(1), pp. 3–9.

[3] Ackoff, R. L. (1999). Ackoff's Best. New York: John Wiley & Sons, pp 170 – 172.

[4] Alcácer, V. & Cruz-Machado, V. (2019). Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems. Engineering Science and Technology, an International Journal, https://doi.org/10.1016/j.jestch.2019.01.006.

[5] Amin, A., Shah, B., Khattak, A. M., Baker, T., & Anwar, S. (2018). Just-in-time Customer Churn Prediction: With and Without Data Transformation. In *2018 IEEE Congress on Evolutionary Computation (CEC)* pp. 1–6. IEEE.

[6] ANSI/DMSC QIF 3.0 (2018). Quality Information Framework (QIF) – An Integrated Model for Manufacturing Quality Information. Digital Metrology Standards Consortium, Inc. (DMSC). https://qifstandards.org/download/ (visited on 16.9.2019).

[7] Autodesk (2013). Autodesk® AutoCAD® 2014, DXF Reference. http://images.autodesk.com/adsk/files/autocad_2014_pdf_dxf_reference_enu.pdf (visited on 16.10.2019).

[8] Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., Mané, D. (2016). Concrete Problems in AI Safety. Retrieved from https://arxiv.org/pdf/1606.06565.pdf

[9] Antonoglou, I., Fidjeland, A. K., Wierstra, D., King, H., Bellemare, M. G., Legg, S., Mnih, V. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.

[10] Barkmeyer E.J. (1996). SIMA Reference Architecture – Part 1: Activity Models, NIST Interagency/Internal Report (NISTIR) 5939. National Institute of Standards and Technology, Gaithersburg, MD, 1996. https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir5939.pdf

[11] Barkmeyer E. and Wallace E.K. (2016). NIST Advanced Manufacturing Series 300-1. Reference Architecture for Smart Manufacturing Part 1: Functional Models. National Institute of Standards and Technology, Gaithersburg, MD, 2016. http://dx.doi.org/10.6028/NIST.AMS.300-1

[12] BDVA (2018). Big Data Challenges in Smart Manufacturing – BDVA white paper: http://www.bdva.eu/sites/default/files/BDVA_SMI_Discussion_Paper_Web_Version.pdf (visited on 6.6.2019).

[13] Bejtlich, R. (2005). The Tao of network security monitoring : beyond intrusion detection. Addison-Wesley.

[14] BDVA (2019). Towards a European Data Sharing Space - Enabling data exchange and unlocking AI potential. BDVA Position Paper. Available at: http://bdva.eu/AIPPP-Vision-paper-PressRelease.

[15] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, Vol. 284(5), pp.28–37.

[16] Blecker, T. Friedrich, G., Kaluza, B., Abdelkafi, N., Kreutler, G. Information and Management Systems for Product Customization. Springer Science + Business Media, Inc. 2005. https://doi.org/10.1007/b101300 (visited on 22.10.2019).

[17] Bruseker, G., Carboni, N., & Guillem, A. (2017). Cultural heritage data management: the role of formal ontology and CIDOC CRM. In *Heritage and Archaeology in the Digital Age* (pp. 93–131). Springer, Cham.

[18] Cai, L. and Zhu, Y. (2015) The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal, Vol. 14, pp. 2. DOI: http://doi.org/10.5334/dsj-2015-002 (visited on 6.6.2019).

[19] Carrasco, J., Cubo, J., & Pimentel, E. (2014). Towards a flexible deployment of multi-cloud applications based on TOSCA and CAMP. In *European Conference on Service-Oriented and Cloud Computing* pp. 278–286. Springer, Cham.

[20] Cavalcante, E., Pereira, J., Alves, M. P., Maia, P., Moura, R., Batista, T. & Pires, P. F. (2016). On the interplay of Internet of Things and Cloud Computing: A systematic mapping study. *Computer Communications*, Vol.89, pp.17–33.

[21] Chandrasekaran, B., Johnson, T.R., & Smith, J.W. (1992). Task-Structure Analysis for Knowledge Modelling. *Communications of the ACM*, Vol. 35(9), pp. 124–137.

[22] Chauhan, S. S., Pilli, E. S., Joshi, R. C., Singh, G., & Govil, M. C. (2018). Brokering in interconnected cloud computing environments: A survey. *Journal of Parallel and Distributed Computing.* DOI; https://doi.org/10.1016/j.jpdc.2018.08.001 (visited on 6.6.2019).

[23] Cheng, S., Li, Y., Tian, Z., Cheng, W., & Cheng, X. (2019). A model for integrating heterogeneous sensory data in IoT systems. *Computer Networks*, Vol. 150, pp. 1–14.

[24] Chen, M., Mao, S., and Liu, Y. (2014). Big data: a survey. *Mobile Networks and Applications*, Vol. 19(2), pp. 171–209.

[25] Datta, S. K., Bonnet, C., Da Costa, R. P. F., & Härri, J. (2016). Datatweet: An architecture enabling data-centric iot services. In *2016 IEEE Region 10 Symposium (TENSYMP)* pp. 343–348. IEEE.

[26] Davies, A. (2004). “Moving base into high-value integrated solutions: a value stream approach”, Industrial and Corporate Change. 13(5), pp. 727–756.

[27] Dijcks, J. P. (2013). “Oracle: Big Data for the Enterprise”, An Oracle Whitepaper, June 2013.

[28] DIN e.V. (2018). DIN/DKE – Roadmap, German standardization roadmap, Industrie 4.0, Version 3. https://www.din.de/blob/65354/57218767bd6da1927b181b9f2a0d5b39/roadmap-i4-0-e-data.pdf (visited on 19.9.2019).

[29] Druzin, B. (2016). Towards a Theory of Spontaneous Legal Standardization. *Journal of International Dispute Settlement*, Vol. 8(3), pp. 403–431.

[30] ElMaraghy, H. (2019). “Smart changeable manufacturing systems”, Procedia Manufacturing, Vol. 28, pp. 3–9 ISSN 2351–9789, DOI: https://doi.org/10.1016/j.promfg.2018.12.002. Available at https://www.sciencedirect.com/science/article/pii/S2351978918313441 (visited 6.6.2019).

[31] EN17007 (2017). Maintenance process and associated indicators.

[32] Eppler, M. (2006). Managing Information Quality: Increasing the Value of Information in knowledge-intensive Products and Processes. 2nd ed. Heidelberg: Springer, 2006. DOI:10.1007/3-540-32225-6

[33] EPSC (2017). *Enter the Data Economy: EU Policies for a Thriving Data Ecosystem*. *EPSC Strategic Notes*, Available at: https://euagenda.eu/upload/publications/untitled-88612-ea.pdf (visited 18.9.2019).

[34] Fahad, M. (2008). Er2owl: Generating owl ontology from er diagram. *In International Conference on Intelligent Information Processing* pp. 28–37, Springer, Boston, MA

[35] Tao, F., Qi, Q., Kusiak, A. (2018). Data-driven smart manufacturing. Journal of Manufacturing Systems. Vol. 48,157–169 https://doi.org/10.1016/j.jmsy.2018.01.006

[36] Fleckenstein M. & Fellows L. (2018). Physical Asset Management vs. Data Management. In: Modern Data Strategy. Springer, Cham.

[37] Freund, J., Jones, J. (2014). Measuring and managing information risk : a FAIR approach. Butterworth-Heinemann.

[38] Frost & Sullivan (2017) The Dawn of Artificial Intelligence—Foreseeing Manufacturing in the Cognitive Era — Foreseeing Manufacturing in the Cognitive Era.

[39] Gao, Z. Y., Liang, Y. Q., & Qiao, S. H. (2016). Relational Database Ontology Discovery Method Based on Formal Concept Analysis. In *3rd Annual International Conference on Mechanics and Mechanical Engineering* (MME 2016). Atlantis Press.

[40] Giustozzia, F. et al. (2018). Context Modeling for Industry 4.0: an Ontology-Based Proposal. In Proc. of International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2018, 3–5 September 2018, Belgrade, Serbia

[41] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, Vol. 5(2), pp. 199–220.

[42] Harlou, U. (2006). Developing product families based on architectures: Contribution to a theory of product families. Kgs. Lyngby: Technical University of Denmark.

[43] Hedberg, T., Lubell, J., Fischer, L., Maggiano, L. (2016). Testing the Digital Thread in Support of Model-Based Manufacturing and Inspection. *J. Comput. Inf. Sci. Eng*. Vol.16, No 2. https://doi.org/10.1115/1.4032697

[44] Hellström, M., Lagström, L. Selecting the best – smarter purchasing through software at Rolls-Royce, In: REBUS – Towards Relational Business Practices, Final Report. Ed. by Valkokari, K., DIMECC Publications Series No. 14. 2017 pp. 98–104

[45] Helu, M., Hedberg, T., Barnard Feeney, A. (2017). Reference architecture to integrate heterogeneous manufacturing systems for the digital thread. *CIRP Journal of Manufacturing Science and Technology*. Vol. 19. pp. 191–195. https://doi.org/10.1016/j.cirpj.2017.04.002

[46] Herron, J., Gelotte, R. A QIF Case Study – Maintaining the Digital Thread from OEM to Supplier. In the Proc. of the 10th Model-Based Enterprise Summit (MBE 2019), Gaithersburg, Maryland, USA, April 2–4, 2019. https://www.nist.gov/publications/proceedings-10th-model-based-enterprise-summit-mbe-2019 (visited on 5.9.2019).

[47] Hirz, M., Dietrich, W., Gfrerrer, A., Lang, J. (2013). Integrated Computer-Aided Design in Automotive Development: Development Processes, Geometric Fundamentals, Methods of CAD, Knowledge-Based Engineering Data Management. Springer-Verlag, 466 pp. DOI 10.1007/978-3-642-11940-8

[48] Hou, L.B. et al. (2017). Applications of artificial intelligence in intelligent manufacturing: a review. Frontiers of Information Technology & Electronic Engineering, Vol. 18(1), pp.86–98

[49] IEC 2014. IEC 60300-1: Dependability management – Part 1: Guidance for management and application. International Electrotechnical Commission, Geneva

[50] International Data Spaces Association (2018). IDS Reference Architecture Model Industrial Data Space. DOI: http://doi.org/10.13140/RG.2.2.17352.11529 (visited on 2.7.2019).

[51] ISO 10303-1 (1994). Industrial automation systems and integration – Product data representation and exchange – Part 1: Overview and fundamental principles. Standard.

[52] ISO 10303-11 (2004). Industrial automation systems and integration – Product data representation and exchange – Part 11: Description methods: The EXPRESS language reference manual. Standard.

[53] ISO 10303-28 (2007). Industrial automation systems and integration – Product data representation and exchange – Part 28: Implementation methods: XML representations of EXPRESS schemas and data, using XML schemas. Standard.

[54] ISO 55000: 2014. Asset .management – overview, principles and terminology.

[55] ISO 55001: 2014. Asset management – Management systems – Requirements.

[56] ISO/IEC 27000: 2009. Information Technology – Security Techniques – Information Security Management Systems – Overview and Vocabulary. ISO/IEC.

[57] ISO/IEC 9075-1: 2016. Information technology – Database languages – SQL – Part 1: Framework (SQL/Framework). ISO/IEC.

[58] ISO/IEC/IEEE 15288 (2015). Systems and software engineering – System life cycle processes. ISO/IEC/IEEE.

[59] ISO/TC 184/SC 4, a list of STEP standard parts (2004). Web page: https://web.archive.org/web/20041030171929/http://www.tc184-sc4.org/titles/STEP_Titles.htm (visited on 14.10.2019).

[60] ISO/TR 23087: 2018 Automation Systems and integration – the Big Picture of standards.

[61] Jiang, L., Da Xu, L., Cai, H., Jiang, Z., Bu, F., & Xu, B. (2014). An IoT-oriented data storage framework in cloud computing platform. *IEEE Transactions on Industrial Informatics*, Vol. 10(2), pp. 1443–1451.

[62] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260.

[63] Jokinen, L., Vainio, V., & Pulkkinen, A. (2017). *Engineering Change Management Data Analysis from the Perspective of Information Quality*, In: Procedia Manufacturing. 11, p. 1626–1633 8 p. https://doi.org/10.1016/j.promfg.2017.07.312 (visited on 22.10.2019).

[64] Kagermann, H.; Wahlster, W. and Helbig, J. (2013). Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0. Berlin: Industrie 4.0 Working Group of Acatech. https://www.acatech.de/Publikation/recommendations-for-implementing-the-strategic-initiative-industrie-4-0-final-report-of-the-industrie-4-0-working-group/ (visited on 6.6.2019).

[65] Karvonen H., Heikkilä E., Wahlström M. (2019). Artificial Intelligence Awareness in Work Environments. In: Barricelli B. et al. (eds.) Human Work Interaction Design. Designing Engaging Automation. HWID 2018. IFIP Advances in Information and Communication Technology, vol 544. Springer.

[66] Kaisler, S., Armour, F., Alberto Espinosa, J., Money, W. (2013). "Big Data-Issues and Challenges Moving Forward," IEEE, 2013 46th Hawaii International Conf. System Sciences.

[67] Kaufmann, M. (2019). Big Data Management Canvas: A Reference Model for Value Creation from Data. Big Data and Cognitive Computing, Vol. 3(1), pp. 19 DOI: https://doi.org/10.3390/bdcc3010019 (visited on 2.7.2019).

[68] Klemke, T., Nyhuis, P. Lean Changeability – Evaluation and Design of Lean and Transformable Factories. In: World Academy of Science, Engineering and Technology International Journal of Economics and Management Engineering Vol:3, No:5, 2009, pp. 454–461, https://publications.waset.org/157/pdf (visited on 22.10.2019).

[69] Kiritsis, D., Bufardi, A., Xirouchakis, P. (2003). Research issues on product lifecycle management and information tracking using smart embedded systems. Advanced Engineering Informatics, Volume 17, Issues 3–4, July–October 2003, pp. 189–202, https://doi.org/10.1016/j.aei.2004.09.005 (visited on 10.10.2019).

[70] Kortelainen, H., Kunttu, S., Valkokari, P., Ahonen, T., Kinnunen, S.-K., Ali-Marttila, M., Herala, A., and Marttonen-Arola, S. (2015). D2BK Data to Business Knowledge Model. Data Sources and Decision making needs. Fimecc S4Fleet Project 3 SP 1 Fleet information network and decision making. Deliverable 1. 29p.

[71] Kortelainen, H., Hanski, J., Kunttu, S., Kinnunen, S.-K. and Marttonen-Arola, S. (2017). *Fleet service creation in business ecosystems – from data to decisions*. *VTT Technology 309*.

[72] Kortelainen, H., Komonen, K. (2017). Asset management in transformation - the role of international standards in customer-supplier collaboration. In: Martinsuo, M. and Kärri, T. (eds.) Teollinen internet uudistaa palveluliiketoimintaa ja kunnossapitoa. Kunnossapitoyhdistys Promaint ry. Helsinki 2017, pp. 202–2013. (in Finnish).

[73] Kortelainen, H., Happonen, A. Hanski, J. (2019a). From asset provider to knowledge company – Transformation in the digital era. In: *Asset Intelligence through Integration and Interoperability and Contemporary Vibration Engineering Technologies.* Mathew, J., Lim, C. W., Ma, L., Sands, D., Cholette, M. E. & Borghesani, P. (eds.). Springer. p. 333–341. 9 p. (Lecture Notes in Mechanical Engineering).

[74] Kortelainen, H., Saari, L., Valkokari, K., Federley, M., Heilala, J., Huusko, J., & Viljamaa, E. (2019b). Beyond IoT Business. VTT Technical Research Centre of Finland. VTT White Paper https://doi.org/10.32040/WhitePaper.2019.BeyondIoT.

[75] Koutroumpis, P. and Leiponen, A. (2013). Understanding the value of (big) data. 2013 IEEE International Conference on Big Data, Pp. 38–42.

[76] Kunttu, S., Ahonen, T. & Kortelainen, H. (2017). Tiedon jalostusastetta nostaen parempia palveluita ja viisaampia päätöksiä. In: Martinsuo, M. and Kärri, T. (eds.) Teollinen internet uudistaa palveluliiketoimintaa ja kunnossapitoa. Kunnossapitoyhdistys Promaint ry. Helsinki 2017, pp. 15–25. (in Finnish).

[77] Laitila, M. Data monetization: Utilizing data as an asset to generate new revenues for firms. Master's thesis. Aalto University. 2017. Espoo. https://aaltodoc.aalto.fi/bitstream/handle/123456789/28934/master_Laitila_Miikka_2017.pdf?sequence=1&isAllowed=y.

[78] Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A. T. (2010). A survey of mobile phone sensing. *IEEE Communications magazine*, Vol. 48(9), pp. 140–150.

[79] Laney, D., Faria, M., and Duncan, A. D. (2015). Seven Steps to Monetizing Your Information Assets. Technical Report Gartner.

[80] Lee, J., Bagheri, B., Kao, H-A. (2015). Manufacturing Letters, 3. pp18–23.

[81] Lee, J., Lapira, E., Bagheri, B., Kao, H-A. (2013). Recent advances and trends in predictive manufacturing systems in big data environment, Manufacturing Letters. 1. pp 38–41. doi:10.1016/j.mfglet.2013.09.005.

[82] Lee, J. H., Shin, J., & Realff, M. J. (2018). Machine learning: Overview of the recent progresses and implications for the process systems engineering field. Computers & Chemical Engineering.

[83] Leikas, J., Koivisto, R., & Gotcheva, N. (2019). Ethical Framework for Designing Autonomous Intelligent Systems. Journal of Open Innovation: Technology, Market, and Complexity, 5(1), 18. https://doi.org/10.3390/joitmc5010018.

[84] Li, B., Hou, B., Yu, W. et al. (2017). Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering*, Vol. 18 pp. 86–98.

[85] Malik, K. R., Ahmad, T., Farhan, M., Aslam, M., Jabbar, S., Khalid, S., & Kim, M. (2016). Big-data: transformation from heterogeneous data to semantically-enriched simplified data. *Multimedia Tools and Applications*, Vol. 75(20), pp. 12727–12747.

[86] Marr, B. (2019). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. The website of Forbes. https://www.forbes.com/sites/bernard-marr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#6eb02fe160ba (visited on 16.9.2019).

[87] McKinsey. (2017). Smartening up with Artificial Intelligence (AI) - What's in it for Germany and its Industrial Sector? *Digital McKinsey*, 9. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Industries/Semiconductors/Our Insights/Smartening up with artificial intelligence/Smartening-up-with-artificial-intelligence.ashx str 9.

[88] Meariam, L. (2017). Why Intel is buying car-vision company Mobileye for $15.3B Computerworld.13 MARCH 2017 https://www.computerworld.com/article/3180164/why-intel-is-buying-car-vision-company-mobileye-for-153b.html.

[89] MIMOSA (2019). Open Standards for Physical Asset Management. Web site: https://www.mimosa.org/ (visited on 4.7.2019).

[90] MITRE ATT&CK™ [WWW Document], 2019 URL https://attack.mitre.org/ (accessed 9.23.19).

[91] Mohamed, M. (2018). Challenges and Benefits of Industry 4.0: An overview. *International Journal of Supply and Operations Management*, Vol. 5(3) pp. 256–265.

[92] Mortensen, H.H., Harlou, U, Haug, A. Improving Decision Making in the Early Phases of Configuration Projects. International Journal of Industrial Engineering: Theory, Applications and Practice, [S.l.], v. 15, n. 2, p. 185–194, Oct. 2008. ISSN 1943-670X. Available at: <http://journals.sfu.ca/ijietap/index.php/ijie/article/view/119>. (visited on 22.10.2019).

[93] Nicolescu R., Radanliev, R., De Roure, D. (2018). State of the art in IoT - Beyond economic value. IoTUK. https://iotuk.org.uk/wp-content/uploads/2018/08/State-of-the-Art-in-IoT-%E2%80%93-Beyond-Economic-Value2.pdf.

[94] Nilsson, N. J. (2011). The Quest for Artificial Intelligence: A History of Ideas and Achievements. Cambridge University Press.

[95] Niskanen, I., Kantorovitch, J., & Golenzer, J. (2012). Monitoring and Visualization Approach for Collaboration Production Line Environments: A Case Study in Aircraft Assembly. *International Journal on Human Computer Interaction*, Vol. 3(2), pp. 35–50.

[96] NIST (2012). SP800-30 Guide for Conducting Risk Assessments, NIST Special publication.

[97] NIST (2018). Product Definitions for Smart Manufacturing project https://www.nist.gov/programs-projects/product-definitions-smart-manufacturing

[98] Nonaka, I. & Takeuchi, I. (1995). The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation, Oxford University press.

[99] Ochs, T., & Riemann, U. A. (2018). IT Strategy Follows Digitalization. In *Encyclopedia of Information Science and Technology*, Fourth Edition pp. 873–887. IGI Global.

[100] OECD (2017). The Next Production Revolution. The Next Production Revolution. doi:10.1787/f69a68e9-en.

[101] Olesen, J.D. (1998) Pathways to Agility: Mass Customization in Action. National Association of Manufacturers, 1st Edition, 263 pp. https://www.amazon.com/Pathways-Agility-Customization-Association-Manufacturers/dp/0471191752.

[102] Opara-Martins, J., Sahandi, R., & Tian, F. (2016). Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective. *Journal of Cloud Computing*, Vol. 5(1).

[103] Perttula A. Challenges and improvements of verification and validation activities in high volume electronics product development. Tampere: Tampere University of Technology, 2007.

[104] Ping, P., Hermjakob, H., Polson, J. S., Benos, P. V., & Wang, W. (2018). Biomedical informatics on the cloud: a treasure hunt for advancing cardiovascular medicine. *Circulation research*, Vol. 122(9), pp. 1290–1301.

[105] Pulkkinen A, Leino S-P, Papinniemi J. (2017). Transforming ETO Businesses with Enhanced PLM Capabilities. Procedia Manufacturing 11. pp. 1642 – 1650, doi: 10.1016/j.promfg.2017.07.315.

[106] Pulkkinen, A., Anttila, J-P., Leino, S-P. (2019). Assessing the maturity and benefits of digital extended enterprise. Presented in: 29th International Conference on Flexible Automation and Intelligent Manufacturing, University of Limerick, Limerick, Ireland. 2019. To be published in Procedia Manufacturing. 10 pp.

[107] Li, Q., Tang, Q., Chan, I., Wei, H., Pu, Y., Jiang, H., Li, J., Zhou. J. (2018). Smart manufacturing standardization: Architectures, reference models and standards framework, Computers in Industry, Volume 101, 2018, Pages 91–106, ISSN 0166-3615, https://doi.org/10.1016/j.compind.2018.06.005 .

[108] Quintana, V., Rivest, L., Pellerin, P., Kheddouci, F. Will Model-based Definition replace engineering drawings throughout the product lifecycle? A global perspective from aerospace industry. Computers in Industry, 2010. https://doi.org/10.1016/j.compind.2010.01.005 (visited on 1.4.2019).

[109] Quintana, V., Rivest, L., Pellerin, P., Kheddouci, F. Re-engineering the Engineering Change Management process for a drawing-less environment, Computers in Industry, 2012. https://doi.org/10.1016/j.compind.2011.10.003 (visited on 10.10.2019).

[110] Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.,* Vol. 23(4), pp. 3–13.

[111] Reifsnider, K., Majumdar, P. (2013). Multiphysics Stimulated Simulation Digital Twin Methods for Fleet Management, in: 54th AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf., 2013: p. 1578. doi:10.2514/6.2013–1578.

[112] Richter, G., & Scmitz, C. (2019). AI in production: A game changer for manufacturers with heavy assets. McKinsey. Retrieved October 2, 2019, from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ai-in-production-a-game-changer-for-manufacturers-with-heavy-assets#.

[113] Ristoski, P., & Paulheim, H. (2016). Semantic Web in data mining and knowledge discovery: A comprehensive survey. *Web semantics: science, services and agents on the World Wide Web*, Vol. 36, pp. 1–22.

[114] Rosati, J., Ristoski, P., Di Noia, T., Leone, R. D., & Paulheim, H. (2016). RDF graph embeddings for content-based recommender systems. In *CEUR workshop proceedings* Vol. 1673, pp. 23–30. RWTH.

[115] Rowley, J. (2006). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, Vol. 33, No. 2, pp. 163–180.

[116] Sabou, M., Kantorovitch, J., Nikolov, A., Tokmakoff, A., Zhou, X., & Motta, E. (2009). Position paper on realizing smart products: Challenges for semantic web technologies. Paper presented at the International Conference on Semantic Sensor Networks pp. 135–147, USA: ACM Press.

[117] Savola R., Fruwirth C. & Pietikainen A. (2012). Risk-Driven Security Metrics in Agile Software Development – An Industrial Pilot Study, *Journal of Universal computer Science*, Vol. 18(12), pp. 1679–1702. https://doi.org/10.3217/JUCS-018-12-1679

[118] Savola, R.M. (2009). A Security Metrics Taxonomization Model for Software-Intensive Systems. J. Inf. Process. Syst. 5, 197–206. https://doi.org/10.3745/jips.2009.5.4.197.

[119] Schabus, S. & Scholz, J. (2017). Spatially-Linked Manufacturing Data to Support Data Analysis. *Journal for Geographic Information Science*, Vol. 1(15), pp. 126–140.

[120] Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D. and Tufano, P. (2012). "Analytics : The real world use of big data," IBM Global Business Services-Business Analytics and Optimization, Executive Report, IBM Institute for Business Value in collaboration with Saïd Business School at the University of Oxford, October 2012.

[121] Seliger, G. (ed.) (2007). Sustainability in manufacturing. Recovery of Resources in Product and Material Cycles. Springer Verlag.

[122] Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., Wang, L. (2010). DRAFT Modeling, Simulation, Information Technology & Processing Roadmap. Technology Area 11. www.nasa.gov/sites/default/files/atoms/files/2015_nasa_technology_roadmaps_ta_11_modeling_simulation_final.pdf (visited 10.9.2019).

[123] Shingo, S. (1989). Study of Toyota Production System, Edited by Andrew P. Dillon, 304 pp. https://doi.org/10.4324/9781315136509 (visited on 23.10.2019).

[124] Shostack, A. (2008). Experiences threat modeling at Microsoft. CEUR Workshop Proc. 413, 1–11.

[125] Shostack, A. (2014). Threat modeling: designing for security. Wiley.

[126] Seppälä, T., Juhanko, J. and Mattila, J., 2018. Data Ownership and Governance.

[127] Silvola, R. One product data for integrated business processes. Ph. D. Dissertation, University of Oulu. 2018 http://urn.fi/urn:isbn:9789526221144 (Visited on 15.9.2019).

[128] Singh, B. (2017). Connecting IT with Operational and Engineering Technology for Asset Performance Modeling. https://www.bimcommunity.com/news/load/461/connecting-it-with-operational-and-engineering-technology-for-asset-performance-modeling. (Visited on 14.10.2019).

[129] Spijker, A. v. (2014). The New Oil: Using Innovative Business Models to turn *Data Into Profit*. Technics Publications.

[130] Soininen, T., Tiihonen, J., Männistö, T., & Sulonen, R. (1998). Towards a general ontology of configuration. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 12(4), 357–372. doi:10.1017/S0890060498124083 (visited on 20.10.2019).

[131] Tuegel, E. (2012). The Airframe Digital Twin: Some Challenges to Realization, in: 53rd AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf., p. 1812. doi:10.2514/6.2012–1812.

[132] US PRO (2006). U.S. Product Data Association. Initial Graphics Exchange Specification IGES 5.3. https://web.archive.org/web/20120821190122/http://www.uspro.org/documents/IGES5-3_forDownload.pdf (visited on 16.10.2019).

[133] Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review*, Vol. 11(2), pp. 93–136.

[134] de Vallejo, L., I., Scerri, S., Tuikka, T. (eds.) (2019). Towards a European Data Sharing Space. Brussels. BDVA Position paper, April 2019.

[135] Wang, J., Ma, Y., Zhang, L., Gao, R.H., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications, *Journal of Manufacturing Systems*, Vol.48 Part C, pp. 144–156.

[136] Wixom, B. H. and Ross, J. W. (2017). How to monetize your data. *MIT Sloan* Management Review, 58(3).

[137] Woodhouse, J. (2018). Don't forget human psychology in asset management decisions – it's not all about data and analytics'' Key note speech in WCEAM 2018. The 13th World Congress on Engineering Asset Management. Stavanger, Norway. Sept. 24–27, 2018.

[138] World Manufacturing Forum 2018 report (2018). Available at https://www.worldmanufacturingforum.org/report or direct link to document https://docs.wixstatic.com/ugd/03d390_b6ae0b7ab0da48ca90903b3817be00e6.pdf (visited on 6.6.2019).

[139] The World Wide Web Consortium (W3C) website for the Semantic Web. Available at: https://www.w3.org/standards/semanticweb/ (visited on 6.6.2019).

[140] World Wide Web Consortium (2006). Extensible Markup Language (XML) 1.1 (Second Edition). W3C Recommendation 16 August 2006, edited in place 29 September 2006. https://www.w3.org/TR/xml11/ (visited on 4.7.2019).

[141] World Wide Web Consortium (2012). OWL 2 Web Ontology Language – Document Overview (Second Edition). W3C Recommendation 11 December 2012. https://www.w3.org/TR/owl2-overview/ (visited on 5.7.2019).

[142] World Wide Web Consortium (2014). RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014. https://www.w3.org/TR/rdf11-concepts/ (visited on 5.7.2019).

[143] World Wide Web Consortium (2016). Extensible Markup Language (XML). Web page: https://www.w3.org/XML/ (visited on 4.7.2019).

[144] Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. Production and Manufacturing Research, 4(1), 23–45.

[145] Xu, X. (2012). From cloud computing to cloud manufacturing. *Robotics and computer-integrated manufacturing*, Vol. 28(1), pp. 75–86.

[146] Yan Lu, Katherine C. Morris, Simon P. Frechette (2016).Current Standards Landscape for Smart Manufacturing Systems. NIST Interagency/Internal Report (NISTIR) – 8107. https://nvlpubs.nist.gov/nistpubs/ir/2016/NIST.IR.8107.pdf (visited on 19.9.2019).

[147] Yu, B., MacCallum, K. (1996) Product Structure Based Reason Maintenance for Product Configuration. AAAI Technical Report FS-96-03. Compilation copyright © 1996, AAAI .https://www.aaai.org/Papers/Symposia/Fall/1996/FS-96-03/FS96-03-017.pdf (visited on 20.10.2019).

[148] Zhang, Z., Wu, C., & Cheung, D. W. (2013). A survey on cloud interoperability: taxonomies, standards, and practice. *ACM SIGMETRICS Performance Evaluation Review*, Vol. 40(4), pp. 13–22.

[149] Zrenner, J., Pajam Hassan, A., Otto, B., Marx Gómez, J. (2017). Data Source Taxonomy for Supply Network Structure Visibility, in: Wolfgang Kersten, Thorsten Blecker and Christian M. Ringle (Eds.) Digitalization in Supply Chain Management and Logistics, Proceedings of the Hamburg International Conference of Logistics (HICL) – 23).

9. Appendices

APPEDIX 1: Abbreviations and terminology

AI – Artificial Intelligence.

Asset – is something that has potential or actual value to an organisation. Physical assets include tangible items like equipment, machines and production systems. The asset life cycle stages can be determined by the organisation and be titled appropriate to the organisation's needs. An asset can hold value for one or more organisations over its life (ISO 55000-1).

Big data – The use of data in a size that is larger than what can be utilised in common software applications and common computing hardware. This means the exact size of big data is continuously changing due to progress in computer technology and hardware. Big data is classified with the following characteristics (Chen et al., 2014): Volume (great volume), Variety (various modalities), Velocity (rapid generation), and Value (huge value but very low density). There are also extensions to this characterisation, e.g. veracity (Schroeck et al., 2012), value (Dijcks, 2013) and complexity (Kaisler, 2013).

CAD – Computer-Aided Design.

CAE – Computer-Aided Engineering.

CMMS – Computerised Maintenance Management System.

DCS – Distributed Control System.

DIKW – Data, information, knowledge and wisdom. Data, information, knowledge and wisdom (DIKW) hierarchy is widely accepted as a basic model describing levels of understanding of issues under consideration.

DT – Digital twin.

ERP – Enterprise Resource Planning.

ET – Engineering Technology.

Event – The documentation of the planned or actual occurrence of some phenomenon or the reaching of some milestone.

FMECA - Failure Mode Effects and Criticality Analysis.

Hazop - Hazard and operability study.

Identifier – An attribute of an entity that is used to differentiate that entity from other entities of the same kind.

IIoT – Industrial Internet of Things.

IoT – Internet of Things.

IT – Information Technology.

Master data – Data that is essential for the business or operation and is used as the reference or original data as the main data source.

MES – Manufacturing Execution System.

Metadata – Information about the format and value space that is allowable for a given property attribute.

Modelling – Modelling is an approach to define a conceptual representation of a system, concept or other thing that helps humans to understand the target of modelling based on other, already known concepts. A model can be a drawing, diagram, 3D model, or a physical model of the target. Data modelling is an approach to model data, its representation, and the relations of data elements.

MOM – Manufacturing Operations Management systems.

O&M – Operation and Maintenance.

OT – Operational Technology.

PHM - Prognostics and Health Management.

PLC – Programmable Logic Controller.

SCADA – Supervisory Control And Data Acquisition.

PLM – Product life-cycle Management.

SCM –Supply Chain Management.

Semantic data – Data that contains or has a link to the definition of the meaning of the data and its components. The meaning of the data is usually defined in an ontology that defines the concepts and their relations, and the properties of these. An example of the implementation of the semantic data concept is the Semantic Web and the set of technologies that are developed for it (the W3C website on the Semantic Web).

Semantics – Semantics is a study of meaning. In computer science, “data semantics” and “semantic data” refer to data that contains information about its meaning, preferably in computer interpretable form. Semantic data is used for, e.g., knowledge engineering, in which human knowledge is managed, analysed and reasoned or inferred programmatically. In semantic data management, the definition of concepts and their representation in the selected context is given in an ontology. An ontology is one kind of data model.

Stored data – Data is stored statically and can be reutilised later on.

Streamed data – Data Data originates from its source as continuous series of data values and the data is used on the fly. Data is typically not stored for further use.

Structured data – Data have some predefined structure and meaning, i.e. a data model, and it possibly contains several primitive data types. Data may be structured, e.g., in a relational database system or the structure and meaning can be defined as an ontology (semantic data).

Tacit knowledge –Tacit knowledge refers to non-verbalised action models, skills, ideas and thoughts, which give contents to the verbalised company information. Tacit knowledge can be learned only by experience, and communicated only indirectly, through metaphor and analogy (Nonaka & Takeuchi, 1995).

Taxonomy – Taxonomy is an approach to classify things. Especially in natural sciences, taxonomy is used for classifying, e.g. living organisms. Taxonomy, together with semantic data management, offers a natural means to define the classes and their description, and to classify the data.

Typology – In the context of this report, typology can be seen as an approach to define the meaning, structure and type of data in its numerous forms. In computer science and in programming, we are used to defining data types for the data elements, such as variables in the C programming language:

```
double temperature;
```

In this example, a variable named “temperature” is created with the double data type i.e. double-precision floating point number. A data type can also refer, e.g., to a database definition and the data type of a database table element or the type of a semantic ontology object. The data type is information about how the data element is interpreted, either by a human or, e.g., a computer. In computer science and in computing, the type of data influences how the data is processed in a computer, e.g., how the data is stored in a computer’s memory, how much memory capacity it takes and how the processor processes the data.

Unstructured data – Data does not have a predefined structure and meaning, i.e., a data model, but its meaning can be concluded from the context. Written text is often unstructured and it may contain other primitive data types, such as numeric and time values.

APPENDIX 2: Overview of data required for data security (not a comprehensive list)

- Intra-system scope security related data,
  - System composition, configuration, and state data,
    - There can be multiple (partly or fully) pre-calculated composition/configuration alternatives to cater for different eventualities (say, incident response). Aside the configurations, the data concerning their use cases is needed,
    - It might be necessary (when/if this becomes a practical reality in complex systems) to prove the correctness of configurations. The proofs are to be stored,
  - Security (event) history data,
  - System monitoring data, like logs,
  - Back-up data, both "normal" data and security related,
  - Data for configuring measurements, like metrics to be used, frequency, accuracy, locations, collection, instruments, timing, data volumes to be gathered,
  - Data from security monitoring/measurements, according to metrics,
    - Vulnerability scan data,
    - Statistics,
    - Composition/configuration correctness checks,
    - Security alerts, say, suspicious login attempts,
  - Data gained by analysing monitoring data,
    - System security status data, at several level of granulation,
- Data required by / obtained from external organisations/entities by security service operation,
  - Certificates, keys,
  - Security data about the entities to be known to the system; SW, things or persons, e.g. mapping to groups/roles/rules/policies,
  - Security service configurations (e.g. access control lists, rules, pre-, and post conditions, possibly derived from policies and to be enforced on entities),
  - Distributed Blockchains, for non-reputable transactions between participants, if needed/used,
  - Malware signatures/behaviour patterns (for malware scanning),
  - System behaviour patterns for anomaly detection, often learned,
  - IDS – Intrusion Detection System/Trusted Platform data (secured hashes of the monitored SW),
  - In case of deception techniques, data for creating convincing simulations (in connection of Honeypots and the like),
  - Security Intelligence data,
    - Vulnerability data, vulnerability repository watch,
    - Announcements from Internet security monitoring organisations, national services,
    - Alerts from co-operating organisations,
    - Alerts to co-operating organisations,
- Security management data, process descriptions, standard or not (ISO27000),
  - Security objectives,
    - Security level / maturity target
  - Security requirements,
  - Risk analysis,
  - Methodology-specific data, e.g. best practices, FAIR,
  - Back-up strategy and configuration data,
  - Laws and regulations to be conformed to,
  - Data for proving/fulfilling legal obligations concerning data security (e.g.GDPR),
  - Standards to be followed,

- Security policies,
- Security emergency plans,
- Security decision support data (e.g. for counteractions),
- Data about available countermeasure options and their past/expectable effects,
- Data collected for (potential) forensic purposes,
- Data collected for evaluating security level/maturity, for security management (e.g. for achieving a certificate), and/or for operations development purposes (e.g. lessons learned),
- Development/Improvement plans.